

Министерство образования Республики Беларусь
учреждение образования
«Белорусский государственный университет информатики
и радиоэлектроники»

ТЕХНОЛОГИИ ПЕРЕДАЧИ И ОБРАБОТКИ ИНФОРМАЦИИ

TECHNOLOGIES OF INFORMATION TRANSMISSION AND PROCESSING

МАТЕРИАЛЫ МЕЖДУНАРОДНОГО НАУЧНО-ТЕХНИЧЕСКОГО СЕМИНАРА
(Минск, апрель 2026 г.)

Минск, 2026

УДК 654.1:004.056.5
ББК 32.88+32.971.35-5
Т38

Научный программный комитет:

В. Ю. Цветков, Н. И. Листопад, Б. В. Никульшин, И. С. Азаров, В. А. Вишняков,
О. В. Бойправ, А. В. Гусинский, В. В. Старовойтов, В. К. Конопелько, Ю. В.
Бруттан, О. Дж. М. Аль-Фурайджи, Ван Гуоянь, Д. А. Сикорский, А. С. Солонар,
С.Н. Касанин, С.И. Половения, С.Н. Нефедов, С.Ю. Михневич, В.М. Чертков,
Ци Сесин, Гаофэн Ли

Т38 **Технологии** передачи и обработки информации = Technologies of
Information Transmission and Processing : материалы Международного
научно-технического семинара (Минск, апрель 2026 г.). – Минск : БГУИР,
2026. – 207 с.

Сборник содержит статьи, тематика которых посвящена научно-теоретическим
разработкам в области сетей телекоммуникаций, информационной безопасности,
технологий передачи и обработки информации.

Предназначен для научных сотрудников в области инфокоммуникаций,
преподавателей, аспирантов, магистрантов и студентов технических вузов.

Научное издание

Ответственный за выпуск *В. Ю. Цветков*
Компьютерный дизайн и верстка *В. В. Чепикова*

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.
Ул. П. Бровки, 6, 220013, г. Минск,

© УО «Белорусский государственный университет
информатики и радиоэлектроники», 2026

СОДЕРЖАНИЕ

И.И. ЛЕВОНЕНКО, А.Д. РОБАЧЕВСКИЙ, Г. ВАН, В.Ю. ЦВЕТКОВ ОЦЕНКА ВЛИЯНИЯ ПРЕДОБРАБОТКИ ИЗОБРАЖЕНИЙ НА ПРОИЗВОДИТЕЛЬНОСТЬ И ТОЧНОСТЬ ДЕТЕКЦИИ ОБЪЕКТОВ В SLAMOT- СИСТЕМАХ	7
Ю.С. ЯКОВЛЕВ, Ю.В. БРУТТАН ПОДХОДЫ К ПРОЕКТИРОВАНИЮ СИСТЕМЫ КОНТРОЛЯ КАЧЕСТВА ПОКРЫТИЙ ИЗДЕЛИЙ	16
М.М. ЛУКАШЕВИЧ, А.А. ВОРОНОВ, В.В. ВЕНГЕРЕНКО, М.М. ТАТУР ПРОТОТИПИРОВАНИЕ ПОДСИСТЕМЫ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ ГОЛОВЫ ОТ ФОНА В СОСТАВЕ СИСТЕМЫ ЦЕНТРАЛИЗОВАННОЙ ВИДЕОАНАЛИТИКИ НА БАЗЕ МАШИННОГО ЗРЕНИЯ БПЛА	22
В.А. ВИШНЯКОВ, М.О. КАЦКО МНОГОУРОВНЕВАЯ АРХИТЕКТУРА БЛОКЧЕЙНА ДЛЯ ХРАНЕНИЯ МЕДИЦИНСКИХ ДАННЫХ	31
Т.А. ТУМАНОВ, Л. ВЕЙ, А.Д. ВАСЮТИЧ, Ц. МА, С. ЖЭНЬ, А.В. КУРИЛОВИЧ АЛГОРИТМЫ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ CLC И OPCOSA	37
Д.М. ГОРБУКОВ ПРОЕКТ ЕДИНОЙ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ (EAP) ДЛЯ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ИНФОКОММУНИКАЦИОННОЙ КОМПАНИИ: ПЕРЕХОД К УПРАВЛЕНИЮ, ОСНОВАННОМУ НА ДАННЫХ, В РУП «БЕЛТЕЛЕКОМ»	42
М.Ю. ЛОВЕЦКИЙ СРАВНЕНИЕ АЛГОРИТМОВ ПОДБОРА ВЕСОВОГО КОЭФФИЦИЕНТА ОБЪЕДИНЕНИЯ ИЗОБРАЖЕНИЙ АТОМНО-СИЛОВОЙ МИКРОСКОПИИ	49
Г.С. СМОТРУК, С.Н. ПЕТРОВ, Т.А. ПУЛКО РАЗРАБОТКА ПРАВИЛ КОРРЕЛЯЦИИ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В SIEM-СИСТЕМАХ НА ОСНОВЕ НОРМАТИВНОГО ПЕРЕЧНЯ СОБЫТИЙ	55
И.П. КОБЯК КОНЕЧНЫЕ РАЗНОСТИ В ЗАДАЧЕ СИНТЕЗА ВЕРОЯТНОСТИ ПРОПУСКА ОШИБКИ ПРИ НАБЛЮДЕНИИ ВЕКТОРОВ ПЕРЕХОДОВ	61
Ц. ГО, М.И. ЗОРЬКО, С.В. ЛИТВИНОВ, В.Ю. ЦВЕТКОВ МОДЕЛЬ ОПРЕДЕЛЕНИЯ ДИСТАНЦИИ ДО ОБЪЕКТА В КВАЗИСТАЦИОНАРНОЙ МНОГОРОБОТНОЙ СИСТЕМЕ БЕЗ ОДОМЕТРИИ НА ОСНОВЕ СОГЛАСОВАННОЙ ТРИАНГУЛЯЦИИ	66
А.Н. СОЛОВЬЁВ ЭКСПЕРИМЕНТАЛЬНАЯ СИСТЕМА ДЛЯ ИССЛЕДОВАНИЯ КАНАЛА СВЯЗИ ВИДИМЫМ СВЕТОМ НА БАЗЕ SDR	80
А.В. ЛОМАКО, К.А. ХАДЖИНОВА АВТОМАТИЗАЦИЯ ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЙ ДИАГРАММ В ПРОГРАММНЫЙ КОД	90
Д.Ю. ВОРОНОВИЧ, М.Ю. МАКСАК КОМПЛЕКСНАЯ СИСТЕМА УПРАВЛЕНИЯ БЕСПИЛОТНЫМ АВТОМОБИЛЕМ НА БАЗЕ IOT: ИНТЕГРАЦИЯ ЭНЕРГОПИТАНИЯ И ВИДЕОНАБЛЮДЕНИЯ	98
Е.А. МАСЕЙЧИК, А.А.ПИЛЮШКО ОБЗОР И АНАЛИЗ МИРОВЫХ ТЕНДЕНЦИЙ ПОСТРОЕНИЯ СИСТЕМ ЗАЩИТЫ МЕСТ ПРОКЛАДКИ КАБЕЛЕЙ СВЯЗИ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА	104
QICHENG GUO, M.I. ZORKO, E.G. MAKEICHIK, V.YU. TSVIATKOU METHOD FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING	111

XINYU ZHANG	
SENSOR FUSION FOR AUTONOMOUS VEHICLE ENVIRONMENTAL PERCEPTION: A REPRODUCIBLE COMPARISON	117
Д.М. ГОРБУКОВ	
СОВЕРШЕНСТВОВАНИЕ ОРГАНИЗАЦИОННЫХ МЕХАНИЗМОВ УПРАВЛЕНИЯ РАЗРАБОТКОЙ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ В ИНФОРМАЦИОННО-РАСЧЁТНОМ ЦЕНТРЕ (ИРЦ) РУП «БЕЛТЕЛЕКОМ» НА ОСНОВЕ ТЕОРИИ АКТИВНЫХ СИСТЕМ	123
Е.М. ШЕМЕТОВИЧ, Г.А. ФРОЛОВ, М.И. ЗОРЬКО, В.В. ЧЕПИКОВА, В.Ю. ЦВЕТКОВ	
ФОРМИРОВАНИЕ ЭТАЛОННЫХ ПАР СОВМЕЩЕННЫХ РАЗНОМОДАЛЬНЫХ РАЗНОВРЕМЕННЫХ ИЗОБРАЖЕНИЙ ДЛЯ ОЦЕНКИ КАЧЕСТВА ДЕТЕКТИРОВАНИЯ ЛИНИЙ	136
XINYUE SHAO, QICHENG GUO	
MULTI-ROBOT TASK-PATH COLLABORATIVE OPTIMIZATION	143
А.М. ТИМОФЕЕВ, Д.В. ШЛЯХТУН	
ПРОЦЕДУРА ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ НА ОСНОВЕ МЕТОДА ИЗМЕНЕНИЯ СОСТАВА.....	148
BIN ZHANG	
MULTI SENSOR FUSION FOR MOBILE ROBOT PERCEPTION	151
CHI ZHANG	
DYNAMIC CHANNEL MODELING AND PERFORMANCE ANALYSIS FOR UAV COMMUNICATIONS.....	156
YU LIU	
ANALYSIS OF CYCLE CONSISTENT GENERATIVE ADVERSARIAL NETWORKS FOR UNPAIRED IMAGE TRANSLATION.....	161
QIKAI WANG	
SEMANTIC COMMUNICATION SYSTEM FOR VEHICLE DETECTION	165
SHUO WANG	
PERFORMANCE ANALYSIS OF LS AND LMMSE CHANNEL ESTIMATION FOR 5G NR OFDM SYSTEM.....	168
QICHENG GUO, M.I. ZORKO, GUOYAN WANG, V.YU. TSVIATKOU	
ERROR ANALYSIS MODEL FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING SYSTEM BASED ON FIRST-ORDER ERROR PROPAGATION THEORY.....	174
CAN WANG	
ARTIFICIAL NOISE AIDED PHYSICAL LAYER SECURITY FOR LOW EARTH ORBIT SATELLITE COMMUNICATIONS.....	178
JIA ROU WANG	
DEEP LEARNING-BASED SKELETON EXTRACTION	182
QIYAO YANG, JUN MA, QICHENG GUO	
COMPARISON OF DSO AND ORBSLAM3 IN INDOOR ENVIRONMENTS USING MONOCULAR VISUAL SLAM	186
QINGHAN YU	
UAV-TO-SATELLITE IMAGE GEO-LOCALIZATION VIA DUAL-BRANCH DEEP LEARNING: A COMPARATIVE STUDY OF RESNET-18 AND MLP-MIXER	195

CONTENTS

I.I. LIAVONENKA, A.D. ROBACHEVSKIY, WANG GUOYAN, V.YU. TSVIATKOU EVALUATION OF THE IMPACT OF IMAGE PREPROCESSING ON PERFORMANCE AND ACCURACY OF OBJECT DETECTION IN SLAMOT SYSTEMS	7
IU. S. IAKOVLEV, IU. V. BRUTTAN APPROACHES TO DESIGNING A QUALITY CONTROL SYSTEM FOR COATING PRODUCTS	16
M.M. LUKASHEVICH, A.A. VORONOV, V.V. VENGERENKO, M.M. TATUR PROTOTYPING A HEAD-BACKGROUND SEGMENTATION SUBSYSTEM AS PART OF A CENTRALIZED VIDEO ANALYTICS SYSTEM BASED ON UAV MACHINE VISION	22
U.A. VISHNIAKOU, M.A. KATSKO MULTILEVEL BLOCKCHAIN ARCHITECTURE FOR MEDICAL DATA STORING	31
T. A. TUMANOV, L. WEI, A. D. VASIUTICH, J. MA, X. REN, A.V. KURILOVICH CLC AND OPCOSA: NOISE-RESISTANT CODING ALGORITHMS	37
D.M. HARBUKOU UNIFIED ANALYTICS PLATFORM (UAP) PROJECT FOR THE DIGITAL TRANSFORMATION OF AN INFOCOMMUNICATION COMPANY: TRANSITION TO DATA-DRIVEN MANAGEMENT AT RUP «BELTELECOM»	42
M.YU LAVETSKI COMPARISON OF WEIGHT COEFFICIENT SELECTION ALGORITHMS FOR FUSION OF ATOMIC FORCE MICROSCOPY IMAGES	49
H.S. SMATRUK, S.N. PETROV, T.A. PULKO DEVELOPMENT OF RULES FOR CORRELATION OF INFORMATION SECURITY EVENTS IN SIEM SYSTEMS BASED ON A NORMATIVE LIST OF EVENTS	55
I.P. KOBIK FINITE DIFFERENCES IN THE PROBABILITY SYNTHESIS PROBLEM OF MISSING ERRORS WHEN OBSERVING VECTORS	61
Q. GUO, M. I. ZORKO, S.V. LITVINOV, V.YU. TSVIATKOU MODEL FOR DETERMINING THE DISTANCE TO AN OBJECT IN A QUASI- STATIONARY MULTI-ROBOT SYSTEM WITHOUT ODOMETRY BASED ON A CONSISTENT TRIANGULATION	66
A.M. SALAUYOU EXPERIMENTAL SYSTEM FOR VISIBLE LIGHT COMMUNICATION CHANNEL ANALYSIS BASED ON SDR	80
A.V. LOMAKO, K.A. KHADZHYNAYA AUTOMATION OF DIAGRAM IMAGE CONVERSION TO PROGRAM CODE	90
D.Y. VORONOVICH, M.Y. MAKSAK INTEGRATED IOT-BASED DRIVERLESS VEHICLE CONTROL SYSTEM: POWER SUPPLY AND VIDEO SURVEILLANCE INTEGRATION	98
E.A. MASEYCHIK, A.A. PILYUSHKO REVIEW AND ANALYSIS OF GLOBAL TRENDS IN BUILDING SYSTEMS TO PROTECT COMMUNICATION CABLE INSTALLATION SITES FROM UNAUTHORIZED ACCESS	104
QICHENG GUO, M.I. ZORKO, E.G. MAKEICHIK, V.YU. TSVIATKOU METHOD FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING	111
XINYU ZHANG SENSOR FUSION FOR AUTONOMOUS VEHICLE ENVIRONMENTAL PERCEPTION: A REPRODUCIBLE COMPARISON	117

D.M. HARBUKOU	
IMPROVING ORGANIZATIONAL CONTROL MECHANISMS FOR SOFTWARE DEVELOPMENT IN THE INFORMATION AND SETTLEMENT CENTER (ISC) OF RUE «BELTELECOM» BASED ON ACTIVE SYSTEMS THEORY	123
E.M. SHEMETOVICH, G.A. FROLOV, M.I. ZORKO, V.V. CHEPIKOVA, V.YU. TSVIATKOU	
FORMATION OF REFERENCE PAIRS OF MULTIMODAL MULTI-TEMPORAL IMAGES FOR EVALUATING LINE DETECTION QUALITY	136
XINYUE SHAO, QICHENG GUO	
MULTI-ROBOT TASK-PATH COLLABORATIVE OPTIMIZATION	143
A.M. TIMOFEEV, D.V. SHLIAKHTUN	
PERSONAL DATA DEPERSONALIZATION PROCEDURE BASED ON THE COMPOSITION CHANGE METHOD	148
BIN ZHANG	
MULTI SENSOR FUSION FOR MOBILE ROBOT PERCEPTION	151
CHI ZHANG	
DYNAMIC CHANNEL MODELING AND PERFORMANCE ANALYSIS FOR UAV COMMUNICATIONS.....	156
YU LIU	
ANALYSIS OF CYCLE CONSISTENT GENERATIVE ADVERSARIAL NETWORKS FOR UNPAIRED IMAGE TRANSLATION.....	161
QIKAI WANG	
SEMANTIC COMMUNICATION SYSTEM FOR VEHICLE DETECTION	165
SHUO WANG	
PERFORMANCE ANALYSIS OF LS AND LMMSE CHANNEL ESTIMATION FOR 5G NR OFDM SYSTEM.....	168
QICHENG GUO, M.I. ZORKO, GUOYAN WANG, V.YU. TSVIATKOU	
ERROR ANALYSIS MODEL FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING SYSTEM BASED ON FIRST-ORDER ERROR PROPAGATION THEORY.....	174
CAN WANG	
ARTIFICIAL NOISE AIDED PHYSICAL LAYER SECURITY FOR LOW EARTH ORBIT SATELLITE COMMUNICATIONS.....	178
JIA ROU WANG	
DEEP LEARNING-BASED SKELETON EXTRACTION	182
QIYAO YANG, JUN MA, QICHENG GUO	
COMPARISON OF DSO AND ORBSLAM3 IN INDOOR ENVIRONMENTS USING MONOCULAR VISUAL SLAM	186
QINGHAN YU	
UAV-TO-SATELLITE IMAGE GEO-LOCALIZATION VIA DUAL-BRANCH DEEP LEARNING: A COMPARATIVE STUDY OF RESNET-18 AND MLP-MIXER.....	195

ОЦЕНКА ВЛИЯНИЯ ПРЕДОБРАБОТКИ ИЗОБРАЖЕНИЙ НА ПРОИЗВОДИТЕЛЬНОСТЬ И ТОЧНОСТЬ ДЕТЕКЦИИ ОБЪЕКТОВ В SLAMOT-СИСТЕМАХ

И.И. ЛЕВОНЕНКО^{1,2}, А.Д. РОБАЧЕВСКИЙ¹, Г. ВАН³, В.Ю. ЦВЕТКОВ¹

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

2 – ОАО «Пеленг», Республика Беларусь

3 – Национальный университет оборонных технологий, КНР

Поступила в редакцию 12 апреля 2026

Аннотация. Рассматривается задача экспериментальной оценки влияния методов предобработки изображений на производительность и точность детекции объектов в SLAMOT-системах. С целью выявления оптимальных стратегий улучшения входных данных для мобильных робототехнических платформ, проведено сравнительное исследование эффективности различных алгоритмов предобработки, включая повышение резкости на основе нерезкого маскирования, адаптивную коррекцию контраста CLAHE с гамма-коррекцией, интерполяционное увеличение разрешения и попатчевую обработку (tiling). Экспериментальная оценка выполнена на базе легковесной одностадийной архитектуры YOLOv8n, адаптированной для детекции транспортных средств на аэрофотоснимках посредством программного маппинга классов. В рамках стандартизированного протокола бенчмаркинга осуществляется расчет метрик точности ($mAP@0,5$; Precision; Recall; F1-мера), временных характеристик инференса, а также оценка потребления оперативной и видеопамати. Исследование направлено на определение компромиссов между качеством детекции, вычислительной сложностью и скоростью обработки в условиях ограниченных аппаратных ресурсов бортовых вычислительных систем БЛА.

Ключевые слова: детекция объектов, компьютерное зрение, обработка изображений, SLAMOT.

Введение

В условиях стремительного развития автономных робототехнических систем и беспилотных летательных аппаратов методы компьютерного зрения перестали быть вспомогательным инструментом и превратились в фундаментальный компонент обеспечения восприятия окружающей среды. Ключевую роль в этом процессе играют задачи одновременной локализации, картографирования и трекинга объектов (SLAMOT, Simultaneous Localization and Mapping with Object Tracking), где надежная детекция объектов напрямую влияет на точность построения карты, устойчивость оценки траектории и качество долгосрочного сопровождения динамических целей [1,2]. Однако эффективность нейросетевых детекторов в реальных сценариях существенно зависит не только от архитектуры модели, но и от качества входных визуальных данных, что делает проблему предобработки изображений критически важной для практического развертывания автономных систем.

Использование нейронных сетей в SLAMOT-системах позволяет отказаться от жестко заданных алгоритмических правил и опираться на обученные модели, способные адаптироваться к динамическим, частично наблюдаемым и слабоструктурированным условиям реального мира [3]. Методы предобработки изображений, включая повышение резкости (sharpening), адаптивную коррекцию контраста (CLAHE, Contrast Limited Adaptive Histogram Equalization), гамма-коррекцию и попатчевую обработку (tiling), способны существенно улучшить визуальные признаки, используемые нейронной сетью для принятия решений [4,5]. Однако каждый метод

вносит дополнительные вычислительные затраты и может по-разному влиять на метрики качества детекции, такие как $mAP@0,5$, $Precision$, $Recall$ и $F1$ -мера, а также на показатели производительности: медианное время инференса, потребление оперативной и видеопамати, стабильность кадровой частоты [6].

В настоящей работе проводится систематическая оценка влияния методов предобработки на производительность и точность легковесной архитектуры детекции YOLOv8n [7,8]. Выбор данной модели обусловлен ее широким применением в мобильной робототехнике благодаря оптимальному балансу между скоростью инференса и качеством детекции. Для выбранной архитектуры анализируется изменение метрик точности ($mAP@0,5$; $Precision@0,5$; $Recall@0,5$; $F1$ -мера) и производительности (латентность инференса, использование RAM/VRAM, общее время обработки кадра) при применении четырех конфигураций предобработки: базовой (без обработки), попачковой обработки, комбинации CLAHE с гамма-коррекцией и повышения резкости на основе нерезкого маскирования. Эксперименты выполняются на наборе данных UAVDT (Unmanned Aerial Vehicle Drone Target Detection and Tracking Video Dataset), характерном для задач аэрофотосъемки и мониторинга с БЛА (Беспилотный Летательный Аппарат), что позволяет оценить устойчивость методов в условиях малого размера объектов, изменчивого освещения и динамических сцен [9].

Целью работы является определение возможностей повышения эффективности нейросетевого детектирования объектов в SLAMOT-системах БЛА с помощью различных методов предобработки изображений: повышения резкости, коррекции контраста и попачковой обработки.

Количественные метрики оценки эффективности детекции

Для объективной оценки эффективности детекторов объектов в работе используется стандартный набор метрик, принятый в сообществе компьютерного зрения. В основе всех количественных показателей лежит процедура сопоставления предсказанных ограничивающих рамок с истинными аннотациями с использованием метрики пересечения по объединению (IoU , Intersection over Union), которая вычисляется по формуле

$$IoU = \frac{Area(B_{pred} \cap B_{gt})}{Area(B_{pred} \cup B_{gt})}, \quad (1)$$

где B_{pred} и B_{gt} обозначают предсказанную и истинную ограничивающие рамки соответственно.

Детекция считается корректной, если значение IoU между предсказанной и истинной рамкой превышает порог 0,5 и класс объекта определен верно. В противном случае срабатывание классифицируется как ложное положительное, а пропущенный истинный объект учитывается как ложное отрицательное срабатывание.

Метрика $Precision$ (точность) отражает долю корректных детекций среди всех сработавших детекторов и вычисляется как отношение истинно положительных срабатываний к сумме истинно положительных и ложно положительных

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

где TP (True Positive) – число верно обнаруженных объектов; FP (False Positive) – число ложных детекций.

Высокое значение $Precision$ характеризует способность модели избегать ложных срабатываний, что критически важно для навигационных систем, где ошибочная детекция препятствия может привести к некорректному планированию траектории.

Метрика $Recall$ (полнота) показывает, какая доля истинных объектов была обнаружена моделью, и рассчитывается как отношение истинно положительных срабатываний к сумме истинно положительных и ложно отрицательных

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

где TP (True Positive) – количество корректно обнаруженных объектов; FN (False Negative) – количество пропущенных объектов, которые присутствуют на изображении.

В задачах SLAMOT высокий *Recall* необходим для построения плотной семантической карты и обеспечения надежного долгосрочного трекинга объектов.

Поскольку максимизация *Precision* часто приводит к снижению *Recall* и наоборот, для сводной оценки используется $F1$ -мера – гармоническое среднее между двумя этими величинами, вычисляемое по формуле

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}, \quad (4)$$

Данная метрика достигает максимального значения только при сбалансированно высоких показателях точности и полноты, что делает ее удобной для сравнения моделей в условиях компромисса между типами ошибок.

Ключевой интегральной метрикой в работе выступает $mAP@0,5$ (mean Average Precision при пороге $IoU=0,5$). Ее расчет выполняется путем построения кривой зависимости *Precision* от *Recall* для каждого класса объектов при последовательном изменении порога уверенности модели. Площадь под этой кривой определяет значение Average Precision (AP) для данного класса. В классической интерпретации используется 11-точечная интерполяция

$$AP = \frac{1}{11} \sum_{t \in \{0,0,1,\dots,1,0\}} \max_{\tilde{r} \geq t} p(\tilde{r}), \quad (5)$$

где $p(\tilde{r})$ обозначает максимальное значение *Precision* при уровне полноты не ниже $\tilde{r}$. Итоговое значение mAP получается усреднением AP по всем N категориям объектов в датасете

$$mAP@0,5 = \frac{1}{N} \sum_{i=1}^N AP_i. \quad (6)$$

Порог локализации 0,5 означает, что предсказанная рамка считается верной при перекрытии с истинной не менее чем на 50 %, что представляет собой компромиссное требование, допускающее умеренные погрешности позиционирования в динамических сценах. Значение $mAP@0,5$ чувствительно как к качеству классификации, так и к точности локализации, что делает его универсальным индикатором общей эффективности детектора в составе SLAMOT-системы.

В рамках данного бенчмарка все перечисленные метрики рассчитываются автоматически по результатам инференса на тестовой выборке UAVDT с использованием аннотаций в формате COCO (Common Objects in Context), что обеспечивает воспроизводимость и сопоставимость результатов для различных конфигураций моделей и методов предобработки.

Экспериментальная оценка влияния методов предобработки изображений на производительность и точность детекции объектов в SLAMOT-системах

Экспериментальная оценка эффективности методов предобработки изображений в конвейере SLAMOT выполнена на базе легковесной одностадийной модели детекции YOLOv8n. Выбор данной архитектуры обусловлен ее широким распространением в мобильных робототехнических системах: безжорный механизм детекции в сочетании с оптимизированной архитектурой CSPDarknet (Cross Stage Partial Darknet) обеспечивает высокую скорость инференса при сохранении конкурентоспособной точности распознавания.

Модели семейства YOLO использует веса, предварительно обученные на датасете COCO, и адаптированы для работы с целевыми классами UAVDT (автомобиль, грузовик, автобус) посредством программного маппинга категорий на этапе инференса. Соответствие классов установлено следующим образом: car (COCO класс 2 → UAVDT класс 1), truck (COCO класс 7 →

UAVDT класс 2), bus (COCO класс 5 → UAVDT класс 3). Выбор именно этих категорий обусловлен спецификой задач SLAMOT в городской среде, где транспортные средства являются основными динамическими объектами, требующими отслеживания для построения семантических карт и планирования безопасных траекторий.

Сравнительный анализ осуществляется по совокупности измеряемых характеристик. Время инференса характеризует быстродействие модели при обработке единичного кадра и фиксируется с разделением на этапы предобработки изображения и непосредственного прямого прохода нейронной сети. Потребление памяти отражает пиковый объем оперативной (RAM) и видеопамяти (VRAM), необходимый для выполнения инференса, и измеряется консервативным методом – фиксируется максимальное потребление в течение всего цикла обработки. Размер модели и количество обучаемых параметров определяют требования к хранению весовых коэффициентов и вычислительную сложность архитектуры.

Непосредственное тестирование проводилось в рамках стандартизированного протокола, обеспечивающего воспроизводимость результатов. Порог уверенности устанавливался на уровне 0,2, что ниже стандартного значения 0,5 и позволяет детектировать объекты с умеренной уверенностью модели, что особенно важно для задач SLAMOT, где пропуск объекта может иметь более серьезные последствия, чем ложное срабатывание. Такой выбор порога обусловлен спецификой аэрофотосъемки с БЛА, где объекты часто имеют малый размер, частично перекрыты или находятся в сложных условиях освещения.

Перед замерами выполнялась фаза стабилизации, включавшая 10 итераций прогона модели на репрезентативном входном тензоре, что позволяло минимизировать влияние кэширования, динамической компиляции графа вычислений на итоговые замеры времени.

Характеристики временной эффективности включали медианное время инференса как робастную оценку центральной тенденции, 5-й и 95-й перцентили для оценки диапазона типичных задержек, медианное абсолютное отклонение как меру стабильности производительности, а также минимальное и максимальное время выполнения для выявления экстремальных случаев. Для методов предобработки (повышение резкости, интерполяционное увеличение разрешения в 1,5 раза, адаптивная коррекция контраста CLAHE с гамма-коррекцией, попатчевая обработка) отдельно фиксировалось время выполнения алгоритмов улучшения изображения, что позволяло оценить их вклад в общую задержку обработки кадра.

При использовании метода tiling исходное изображение разбивалось на перекрывающиеся патчи размером 640×640 пикселей, что соответствует стандартному размеру входа моделей семейства YOLOv8, с перекрытием 128 пикселей по каждой оси для предотвращения потери объектов, попадающих на границы разделов. После выполнения детекции на каждом патче результаты со всех фрагментов объединялись в единое пространство координат с применением нежесткого не-максимального подавления с порогом $IoU=0,7$. Выбор повышенного порога подавления дубликатов обусловлен спецификой попатчевой обработки: соседние патчи могут содержать частично перекрывающиеся детекции одного и того же объекта, и стандартный порог 0,5 мог бы привести к излишне агрессивному удалению валидных предсказаний, тогда как значение 0,7 позволяет сохранить детекции частично перекрывающихся объектов, обеспечивая баланс между устранением дубликатов и полнотой обнаружения. Визуализация результатов осуществлялась путем отрисовки только ограничивающих рамок без текстовых подписей классов и значений уверенности, что соответствовало требованиям к чистоте экспериментальной оценки без визуальных артефактов и исключало влияние затрат на рендеринг текста на замеры производительности.

Метод повышения резкости (sharpening), реализованный в данном исследовании, основан на технике нерезкого маскирования. Алгоритм применяет гауссово размытие к исходному изображению с ядром размером 9×9 пикселей и стандартным отклонением 10,0 для получения низкочастотной компоненты. Затем осуществляется взвешенное сложение исходного изображения и размытой версии с коэффициентами 1,5 и -0,5 соответственно. Данная операция эквивалентна добавлению к исходному изображению высокочастотной компоненты, полученной путем вычитания размытой версии из оригинала, что приводит к усилению контраста на границах объектов. Применительно к задаче детекции транспортных средств на аэрофотоснимках UAVDT, метод направлен на компенсацию размытия, вызванного движением

беспилотной платформы и вибрацией камеры, улучшая тем самым видимость мелкогазмерных объектов для сверточной нейронной сети.

На рис. 1 приведены примеры изображений из тестовой выборки UAVDT, использованные для бенчмаркинга нейросетевых моделей.

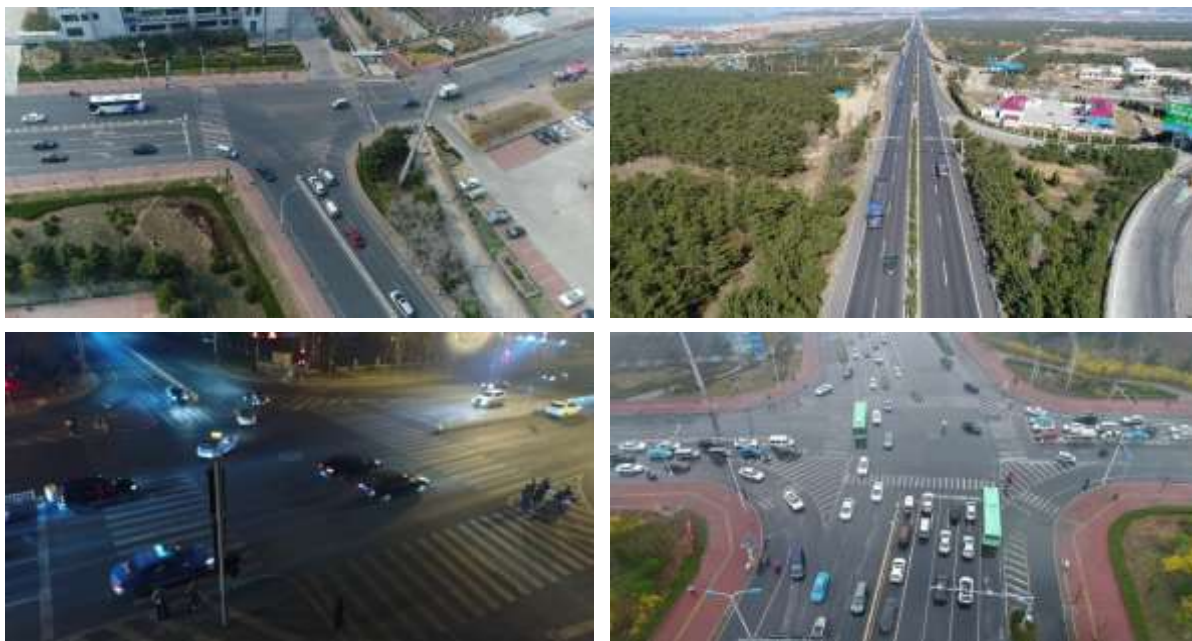


Рис. 1. Примеры кадров из тестовой выборки для бенчмаркинга детекторов

Несмотря на то, что на рис. 1 в качестве примера представлены лишь 4 кадра, они отражают визуальное разнообразие сцен (различные ракурсы, условия освещения, плотность объектов). Экспериментальная оценка выполнялась для 300 изображений, отобранных из полного набора данных UAVDT, содержащего более 4000 кадров.

На рис. 2 представлены результаты детектирования объектов с помощью модели YOLOv8n без предварительной обработки входных изображений.



Рис. 2. Детекция объектов моделью YOLOv8n без предобработки

На рис. 3 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений методом попатчевой обработки.



Рис. 3. Детекция объектов моделью YOLOv8n с предобработкой методом tiling

Как следует из представленных на рис. 3 изображений, применение стратегии попатчевой обработки привело к существенному увеличению количества корректно локализованных объектов, особенно малоразмерных, что выразилось в значительном росте числа истинно положительных срабатываний. Данный эффект имеет критическое значение для задач SLAMOT, где максимальная полнота восприятия динамических объектов является приоритетным требованием для обеспечения безопасности навигации и построения плотной семантической карты.

На рис. 4 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений на основе методов CLAHE и гамма-коррекции.



Рис. 4. Детекция объектов моделью YOLOv8n с предобработкой методами CLAHE и гамма-коррекции

Использование комбинации методов CLANE и гамма-коррекции обусловило общее снижение количества генерируемых детекций. Несмотря на наблюдаемую частичную фильтрацию отдельных типов ложных срабатываний, данный процесс сопровождался возникновением новых ошибок классификации и локализации. Такое неоднозначное влияние свидетельствует о том, что глобальное улучшение контраста не всегда способствует улучшению экстракции признаков нейронной сетью в условиях аэрофотосъемки и может приводить к потере информативных текстурных особенностей, необходимых для надежной детекции.

На рис. 5 представлены результаты детектирования объектов с помощью модели YOLOv8n, с предварительной обработкой входных изображений на основе метода повышения резкости (sharpening).



Рис. 5. Детекция объектов моделью YOLOv8n с предобработкой методом sharpening

Метод повышения резкости на основе нерезкого маскирования продемонстрировал эффективность в сценариях с динамической съемкой, характеризующихся значительным движением камеры и сопутствующим размытием объектов. Применение данного метода позволило компенсировать деградацию высокочастотных компонентов изображения, вызванную вибрацией беспилотной платформы и изменением точки съемки в процессе полета. Усиление контраста на границах объектов способствовало более надежному извлечению признаков сверточной нейронной сетью, что привело к снижению количества частичных ошибок детекции, в частности, ложноположительных срабатываний на текстурированных участках фона и ложноотрицательных результатов для малоконтрастных транспортных средств.

В табл. 1 приведены результаты сравнительной оценки модели YOLOv8n для четырех конфигураций: без предобработки, попатчевая обработка, комбинированный метод CLANE с гамма-коррекцией и sharpening.

Табл. 1. Характеристики детектирования объектов с помощью YOLOv8n

Конфигурация	Среднее время инференса, с	Потребление RAM, МБ	$mAP@0,5$	$Precision@0,5$	$Recall@0,5$	$F1@0,5$
Без предобработки	0,0159	671,77	4,03	57,10	0,17	0,34
Tiling	0,0833	704,35	4,76	50,93	0,29	0,58
CLANE + гамма-коррекция	0,0305	687,36	3,6	55,21	0,14	0,27
Sharpening	0,0177	682,63	4,66	55,27	0,17	0,34

Представленные данные демонстрируют низкие значения $mAP@0,5$ для всех конфигураций, что обусловлено существенным доменным смещением между обучающей выборкой COCO и тестовым датасетом UAVDT. Метод попачечной обработки показал наибольшую эффективность, увеличив $Recall@0,5$ до 0,29 и $F1$ до 0,58 относительно базового уровня, однако ценой пятикратного роста времени инференса. Наблюдаемый дисбаланс между высоким $Precision$ (57–59 %) и низким $Recall$ (0,14–0,29) отражает консервативное поведение YOLOv8n: модель надежно детектирует лишь наиболее очевидные объекты, пропуская малоразмерные транспортные средства, характерные для аэрофотосъемки. Tiling частично компенсирует это ограничение за счет сохранения пространственного разрешения, критичного для детекции мелких объектов.

Применение комбинации методов CLAHE и гамма-коррекции привело к системной деградации всех метрик качества детекции: $mAP@0,5$ снизился на 11 % (с 4,03 до 3,6), $Recall$ упал на 18 % (с 0,17 до 0,14), а интегральная метрика $F1$ показала наибольшее снижение среди всех методов – на 21 % (с 0,34 до 0,27). При этом время инференса увеличилось в 1,9 раза (с 0,0159 с до 0,0305 с), что обусловлено вычислительной сложностью алгоритма адаптивной нормализации гистограммы, требующего локальной обработки каждого тайла изображения размером 8×8 пикселей. Наблюдаемое ухудшение характеристик объясняется фундаментальным несоответствием между принципами работы CLAHE и особенностями предобученной модели YOLOv8n: агрессивное повышение локального контраста, особенно в однородных областях (небо, дорожное покрытие, водная поверхность), приводит к генерации артефактов в виде ложных границ и текстурного шума, которые интерпретируются сверточной сетью как потенциальные объекты детекции, увеличивая количество ложноположительных срабатываний.

Метод повышения резкости (sharpening) продемонстрировал промежуточные результаты: при незначительном увеличении времени инференса на 11 % (с 0,0159 с до 0,0177 с) и минимальном росте потребления памяти (на 1,6 %) удалось достичь прироста $mAP@0,5$ на 16 % относительно базовой конфигурации (с 4,03 до 4,66). При этом метрики $Precision$ (55,27 %) и $Recall$ (0,17) остались на уровне, близком к базовому, что свидетельствует о способности метода sharpening улучшать общую точность детекции без существенного влияния на полноту обнаружения. Умеренное усиление высокочастотных компонентов изображения способствует лучшей локализации границ объектов, компенсируя размытие, вызванное движением камеры, однако не обеспечивает такого прироста $Recall$, как метод tiling, поскольку не решает проблему недостаточного разрешения для малоразмерных объектов.

Заключение

На основе сравнительного бенчмаркинга конфигураций нейросетевого детектора определена оптимальная конфигурация пайплайна предобработки, обеспечивающая максимальную устойчивость детекции при минимальных вычислительных затратах. Показано, что методы предобработки изображений оказывают существенное влияние на производительность и точность детекции объектов в SLAMOT-системах, однако характер этого влияния неоднозначен и зависит от специфики применяемого алгоритма.

Экспериментальные результаты свидетельствуют о низких абсолютных значениях $mAP@0,5$ (3,6–4,76) для всех конфигураций вследствие доменного смещения между COCO и UAVDT, что подтверждает необходимость дообучения модели. Тем не менее, методы предобработки позволяют оптимизировать отдельные аспекты детекции: Tiling обеспечил наилучшую полноту ($Recall@0,5 = 0,29$; $F1 = 0,58$) с приростом 70 %, однако пятикратное увеличение времени инференса ограничивает его применимость в системах реального времени. Sharpening продемонстрировал наиболее сбалансированный результат (прирост $mAP@0,5$ на 16 % при затратах времени +11 %), что делает его перспективным для бортовых платформ, тогда как CLAHE с гамма-коррекцией привел к незначительной системной деградации метрик.

Таким образом, модель YOLOv8 продемонстрировала свою пригодность для интеграции в SLAMOT-системы благодаря высокой скорости инференса (15,9 мс на кадр в базовой конфигурации) и умеренному потреблению памяти, однако для достижения требуемого уровня точности в задачах аэрофотосъемки необходима комбинация различных методов предобработки с последующим дообучением на репрезентативной выборке UAVDT.

Благодарности. Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (БРФФИ) и Национального фонда естественных наук Китая (НФЕНК) в рамках проекта T25КИ-054 «Технология многокамерной локализации и построения карты открытого пространства».

EVALUATION OF THE IMPACT OF IMAGE PREPROCESSING ON PERFORMANCE AND ACCURACY OF OBJECT DETECTION IN SLAMOT SYSTEMS

I.I. LIAVONENKA, A.D. ROBACHEVSKIY, WANG GUOYAN, V.YU. TSVIATKOU

Abstract. This paper presents an experimental evaluation of the impact of image preprocessing methods on the performance and accuracy of object detection in SLAMOT (Simultaneous Localization, Mapping, and Object Tracking) systems. To identify optimal strategies for enhancing input data for mobile robotic platforms, a comparative study was conducted on the effectiveness of various preprocessing algorithms, including unsharp masking-based sharpening, CLAHE (Contrast Limited Adaptive Histogram Equalization) with gamma correction, interpolation-based resolution upscaling, and patch-based processing (tiling). The experimental evaluation was performed using the lightweight single-stage YOLOv8n architecture, adapted for vehicle detection in aerial imagery through programmatic class mapping. Within a standardized benchmarking protocol, accuracy metrics (*mAP@0.5*; *Precision*; *Recall*; *F1-score*), inference timing characteristics, and RAM/VRAM consumption were computed. The study aims to determine trade-offs between detection quality, computational complexity, and processing speed under the hardware resource constraints of onboard computing systems.

Keywords: object detection, computer vision, image processing, SLAMOT.

Список литературы

1. Cadena C., Carlone L., Carrillo H. et al. // IEEE Transactions on Robotics. 2016. Vol. 32, No. 6. P. 1309–1332.
2. Placed J.A., Strader J., Carrillo H. et al. // IEEE Transactions on Robotics. 2023. Vol. 39, No. 5. P. 3403–3438.
3. Bochkovskiy A., Wang C.Y., Liao H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection // arXiv preprint arXiv:2004.10934. 2020.
4. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization // Graphics Gems IV. Academic Press Professional, Inc. 1994. P. 474–485.
5. Gonzalez R.C., Woods R.E. Digital Image Processing. 4th ed. Pearson, 2018.
6. Everingham M., Eslami S.M.A., Van Gool L. et al. // International Journal of Computer Vision. 2015. Vol. 111, No. 1. P. 98–136.
7. Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8 // GitHub Repository. 2023. URL: <https://github.com/ultralytics/ultralytics>
8. Liu W., Anguelov D., Erhan D. et al. SSD: Single Shot MultiBox Detector // Proc. of ECCV. 2016. P. 21–37.
9. Du D., Qi Y., Yu H. et al. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking // arXiv preprint arXiv:1804.00518. 2018.

УДК 621.39:004.932

ПОДХОДЫ К ПРОЕКТИРОВАНИЮ СИСТЕМЫ КОНТРОЛЯ КАЧЕСТВА ПОКРЫТИЙ ИЗДЕЛИЙ

Ю.С. ЯКОВЛЕВ, Ю.В. БРУТТАН

*Псковский государственный университет, Российская Федерация**Поступила в редакцию 6 марта 2026*

Аннотация. В статье рассматриваются подходы к проектированию систем контроля качества покрытий, ориентированных на промышленное применение. Основное внимание уделено неконтактным и неразрушающим методам, включая машинное зрение, оптические, вихретоковые, термографические и спектральные средства контроля. Показано, что эффективность таких систем зависит не только от выбранного метода измерения, но и от того, насколько корректно система встроена в производственный процесс.

Ключевые слова: контроль качества покрытий, неразрушающий контроль, неконтактные методы, машинное зрение, оптическая когерентная томография, терагерцовая диагностика, активная термография, мультисенсорные системы.

Введение

Качество покрытия в промышленном изделии определяет не только его внешний вид, но и ресурс, коррозионную стойкость, износостойкость, тепловую стабильность и надежность в эксплуатации. По этой причине контроль покрытий выступает не отдельной завершающей операцией, а важной частью технологического контура производства. Развитие интеллектуального производства приводит к переходу от эпизодической инспекции к непрерывному и автоматизированному контролю, который по возможности встраивается непосредственно в производственную линию и опирается на методы неразрушающего контроля, машинное зрение и анализ данных [1–5]. Для задач, связанных с поверхностными и подповерхностными дефектами, особое значение приобретают бесконтактные методы: они позволяют обследовать изделия сложной формы, работать на высоких скоростях линии, снижать зависимость результата от оператора и уменьшать риск повреждения контролируемой поверхности [1, 4, 6].

При проектировании системы контроля качества покрытия в первую очередь нужно определить, какие свойства слоя и какие типы дефектов подлежат контролю. В зависимости от назначения покрытия и условий эксплуатации контролируемыми параметрами выступают сплошность слоя, его толщина и равномерность, наличие трещин, пор, кратеров, непрокрасов, наплывов, включений и загрязнений, состояние интерфейса «покрытие-подложка», а также структурные признаки, косвенно связанные с адгезией, шероховатостью и долговечностью [1, 6–9]. Поэтому систему контроля целесообразно рассматривать не как набор отдельных приборов, а как связку датчиков, алгоритмов обработки и средств встраивания в производственный процесс.

Процессы и свойства, относящиеся к контролю качества покрытий

Качество покрытия обычно оценивают по нескольким группам признаков: геометрии слоя, его физико-химическому состоянию и наличию дефектов на поверхности и в зоне интерфейса. Наиболее универсальной характеристикой остается толщина покрытия. Для металлических, защитных, декоративных и функциональных слоев именно толщина во многом определяет

выполнение целевой функции; поэтому в обзорных работах по измерению толщины покрытий она рассматривается как базовый объект контроля [6, 10]. Вместе с тем одного измерения толщины недостаточно. Даже при формальном соблюдении толщины изделие может иметь локальную неравномерность, пористость, нарушения сплошности, скрытые полости, межслойные дефекты или поверхностные изъяны, сформировавшиеся из-за особенностей подготовки подложки и самого процесса нанесения [1, 7, 9].

Не менее важен контроль морфологии поверхности. Для многих покрытий именно текстура, микрорельеф и равномерность распределения материала определяют адгезионные и трибологические свойства. Исследования, посвященные связи шероховатости, сцепления и коррозионной стойкости, показывают, что контроль поверхностного профиля нельзя сводить к одному интегральному параметру; необходим анализ набора признаков, характеризующих форму поверхности и ее статистическую неоднородность [11, 12]. Это особенно важно для покрытий, нанесенных на предварительно обработанные подложки, где микрогеометрия основы влияет на рост дефектов и поведение интерфейса.

Отдельный класс задач связан с распознаванием собственно дефектов покрытия. Для покрытий, формируемых методами распыления, напыления, осаждения из паровой фазы и других высокопроизводительных технологий, характерны такие дефекты, как кратеры, поры, вздутия, трещины, инородные включения, участки недостаточного покрытия и локальные утолщения [1, 8, 13]. Для покрытий вакуумного напыления и тонких пленок отдельно рассматриваются дефекты роста, возникающие вследствие топографии подложки, частиц загрязнений и особенностей роста пленки [8]. С точки зрения проектирования системы контроля это означает, что часть дефектов проявляется как изменение яркостно-текстурной картины на поверхности, а часть требует анализа внутренней структуры или контраста, связанного с электромагнитными, тепловыми либо спектральными свойствами материала.

С учетом сказанного задачи контроля покрытий можно свести к четырем основным направлениям: визуальный контроль поверхности, измерение толщины и равномерности слоя, выявление подповерхностных и межслойных нарушений, а также оценка признаков, связанных с долговечностью покрытия. [1, 6, 10, 11]. От этого набора задач и следует отталкиваться при выборе методов контроля и построении системы.

Неконтактные методы анализа покрытий

На практике чаще всего используют оптические системы машинного зрения. Это связано с высокой скоростью контроля, сравнительно простой интеграцией в производственную линию и возможностью обнаруживать видимые поверхностные дефекты на больших участках. Обзоры, посвященные автоматическому контролю поверхностных дефектов, показывают, что машинное зрение эффективно при поиске царапин, пятен, вмятин, раковин, загрязнений и других нарушений внешнего слоя, особенно если задача формулируется как обнаружение аномалии, классификация дефекта или сегментация дефектной области [2–5]. Для изделий с покрытием оптические методы особенно полезны на участках после нанесения и после сушки, когда требуется быстрый контроль внешнего вида и сплошности поверхности. Однако они ограничены в работе с подповерхностными дефектами и чувствительны к бликам, вариативности освещения, отражательной способности и низкому контрасту дефектов [2, 4, 5].

Для преодоления этих ограничений все большее значение приобретают томографические и спектральные методы. Оптическая когерентная томография (ОСТ) перспективна для послойного анализа покрытий, особенно в тех случаях, когда необходимо оценить внутреннюю структуру прозрачных или полупрозрачных слоев [13–17]. В публикациях по защитным, морским и автомобильным покрытиям показано, что ОСТ способна выявлять толщину слоев, подповерхностные неоднородности и границы раздела без механического контакта с изделием [13, 14, 16]. Для проектирования системы это особенно важно, когда внешний вид поверхности не отражает реального состояния внутренних слоев. При этом ограничения ОСТ связаны с ограниченной глубиной проникновения, зависимостью от оптических свойств материала и необходимостью более сложной калибровки по сравнению с обычной видеосъемкой.

Терагерцовые методы также ориентированы на бесконтактный анализ многослойных структур и измерение толщины диэлектрических и защитных покрытий. Работы последних лет

показывают применимость THz-подходов к оценке толщины термобарьерных покрытий, контролю состояния покрытых стальных конструкций и измерению параметров слоев в задачах промышленного мониторинга [15, 17, 18]. С инженерной точки зрения терагерцовая диагностика ценна тем, что обеспечивает доступ к подповерхностной информации и пригодна для многослойных систем, где обычная оптика дает только картину внешнего слоя. Ограничениями остаются стоимость аппаратуры, чувствительность к материалу и геометрии объекта, а также необходимость сложной интерпретации временных и спектральных сигналов.

Активная термография занимает промежуточное положение между поверхностной и глубинной диагностикой. В этой группе методов объект возбуждается тепловым импульсом или сканирующим источником, а анализ ведется по динамике температурного отклика. Исследования, посвященные контролю толщины и неравномерности покрытий, показывают, что активная термография подходит для крупноформатного и достаточно быстрого обследования, особенно в случаях, когда дефект влияет на теплообмен или тепловую инерцию слоя [19]. Для производственных условий это важно тем, что тепловизионные системы могут охватывать большую область за один цикл наблюдения. Однако метод требует аккуратного управления возбуждением, зависит от эмиссионных свойств поверхности и чувствителен к внешним тепловым помехам.

Вихрековые методы широко применяются для неразрушающего контроля толщины непроводящих покрытий на проводящей подложке и для анализа проводящих слоев при подходящем сочетании материалов. Современные исследования демонстрируют как развитие традиционных схем измерения, так и более сложные конструкции датчиков, снижающие влияние отрыва и обеспечивающие более устойчивое определение толщины [20]. С точки зрения проектирования вихрековые методы привлекательны своей технологической зрелостью и пригодностью к промышленной автоматизации. Вместе с тем они зависят от электромагнитных свойств материала и, как правило, дают ограниченную морфологическую информацию: для распознавания типа дефекта одного такого канала часто недостаточно.

К бесконтактным и близким к бесконтактным физическим методам относятся также рентгенофлуоресцентный анализ и лазерно-спектроскопические подходы. Рентгенофлуоресцентные системы давно применяются для контроля толщины и состава покрытий в поточных линиях, особенно там, где требуется не только оценка геометрии слоя, но и подтверждение его химической природы [21]. Лазерно-индуцированная эмиссионная спектроскопия, в свою очередь, используется для анализа состава и глубинного профилирования слоев, в том числе в режимах, ориентированных на производственный контроль покрытых листовых материалов [22]. Эти методы особенно полезны, когда качество покрытия понимается не только как отсутствие дефектов формы, но и как соответствие состава требуемой технологии. Их ограничение состоит в более сложной аппаратной реализации, требованиях к безопасности и необходимости тщательно согласовывать режим измерения с материалом изделия.

Перспективным направлением выступают мультисенсорные системы. Публикации по комбинированному измерению толщины с использованием, например, хроматического конфокального и вихрекового каналов, а также обзоры по мультисенсорному слиянию данных в умном производстве показывают, что один физический метод редко покрывает весь спектр задач контроля [7]. Визуальный канал хорошо работает с поверхностной морфологией, но хуже – с внутренней структурой; электромагнитный канал чувствителен к толщине и материалу, но почти не интерпретирует форму дефекта; томографический канал информативен, но менее производителен. Поэтому реальная система контроля все чаще проектируется как комбинация взаимодополняющих сенсоров.

Модели и алгоритмы обработки данных

Развитие систем контроля покрытий сопровождалось изменением подходов к обработке данных. В ранних системах машинного зрения чаще всего использовались пороговая обработка, выделение границ, анализ текстур, фильтрация и сравнение с эталоном. Эти методы по-прежнему применяются там, где тип дефекта заранее известен, а условия съемки мало меняются. Их основное достоинство состоит в понятной логике работы и умеренных требованиях к вычислительным ресурсам, что важно для встроенных систем реального времени [3, 4].

По мере усложнения изделий и увеличения разнообразия дефектов всё шире используются методы машинного и глубокого обучения. В таких системах задача контроля обычно сводится к обнаружению дефектов, их классификации, сегментации дефектных областей или поиску аномальных участков [2–5]. Нейросетевые модели хорошо работают в условиях сложного фона и изменчивой геометрии дефектов, однако их внедрение в производство связано с рядом практических трудностей. К ним относятся дисбаланс выборок, высокая стоимость экспертной разметки и ограниченная переносимость обученных моделей между разными типами изделий, покрытий и условиями съемки [5].

По этой причине для проектирования системы важен не только выбор архитектуры нейросети, но и стратегия формирования данных. Актуальны методы *few-shot* и *weakly supervised learning*, а также аномалийный подход, когда модель учится норме, а не каждому классу дефектов по отдельности [1, 4, 5]. Для контроля покрытий это особенно актуально, поскольку многие дефекты редки, а их морфология может сильно варьироваться. Кроме того, для физических методов контроля – например, термографии, терагерцовой или вихретоковой диагностики – алгоритмический уровень включает не только классификатор, но и процедуры реконструкции сигнала, подавления шумов, извлечения информативных признаков, калибровки и слияния каналов [18, 20].

В составе алгоритмического уровня можно выделить несколько функций: реконструкцию измеряемых параметров, интерпретацию полученных данных и формирование производственного решения. На практике они тесно связаны между собой, и качество работы системы зависит от согласованности всех этих этапов. Даже точная модель распознавания не решает задачу, если измерительный процесс нестабилен или, если решение приходит слишком поздно для воздействия на процесс.

В современных системах контроля также важна связь алгоритмического уровня с технологическим контуром производства. Модель должна не только выявлять дефект или оценивать параметр покрытия, но и учитывать контекст измерения: тип изделия, стадию технологического процесса, режим нанесения, историю настройки оборудования и допустимые пределы отклонений. В такой постановке алгоритм становится частью системы поддержки принятия решений, где результат анализа используется для локальной отбраковки, перенастройки режима нанесения, назначения повторного контроля или передачи данных в производственную аналитику.

Архитектурные подходы к проектированию системы

На практике используются несколько схем построения систем контроля покрытий. К ним относятся автономные постконтрольные станции, встроенные средства контроля в линии и мультимодальные системы с иерархическим принятием решения.

Автономная постконтрольная станция размещается после нанесения покрытия и ориентирована на отбраковку или сортировку готовых изделий. Ее сильная сторона – относительная простота интеграции и возможность использовать более сложные сенсоры, поскольку измерение вынесено за пределы критического технологического процесса. Недостаток состоит в том, что система выявляет дефект уже после его возникновения и не всегда позволяет скорректировать процесс в реальном времени [1, 21].

Встроенный контроль предполагает размещение сенсоров непосредственно в линии, а решение принимается синхронно с движением изделия. Такой подход характерен для рулонных, листовых, транспортных и роботизированных процессов нанесения покрытия. Он требует высокой устойчивости к вибрациям, изменению освещенности, загрязнению оптики и ограничениям по времени отклика [1, 6]. Встроенный контроль особенно эффективен тогда, когда результат измерения может использоваться для оперативной коррекции режима распыления, скорости подачи, температуры, расстояния до поверхности или других технологических параметров.

Мультимодальная система с иерархическим принятием решения использует быстрый канал широкого обзора, как правило оптический, для выполнения первичного скрининга, а более дорогие или медленные каналы включаются выборочно для уточняющего анализа подозрительных областей. Подобная архитектура хорошо согласуется с современными

тенденциями интеллектуального производства, где важны и производительность, и глубина анализа [1, 7]. Иерархический подход позволяет уменьшить вычислительную нагрузку и стоимость владения системой, одновременно сохраняя высокую информативность контроля.

По структуре такая система обычно включает сенсорный уровень, средства промышленного сбора данных, вычислительный уровень на периферийных узлах или промышленных ПК и уровень интеграции с остальной производственной системой. При этом данные контроля используются не только для локального решения о годности изделия, но и для анализа причин возникновения дефектов, настройки оборудования и накопления производственной статистики [1].

Ограничения и актуальные направления развития

Основное ограничение носит физический характер: разные типы покрытий и дефектов по-разному взаимодействуют со светом, теплом, электромагнитным полем и различными видами излучения. Поэтому метод, эффективный для прозрачных многослойных покрытий, может оказаться малоинформативным для сильно рассеивающих или проводящих поверхностей [13–20]. Не меньшее значение имеют и производственные условия: загрязнение, вибрации, нестабильная геометрия изделия, высокая скорость линии и изменчивость партий снижают устойчивость решений, которые в лаборатории показывают хорошие результаты.

Отдельные трудности связаны с данными и алгоритмами. Для покрытий характерны редкие дефекты, выраженный дисбаланс классов и ограниченные возможности разметки, поэтому обучение полностью контролируемых моделей оказывается затруднительным [2, 5]. Кроме того, производству требуются не только точность, но и объяснимость: оператору и технологу важно понимать, почему участок признан дефектным и с каким физическим фактором это может быть связано. По этой причине всё больший интерес представляют гибридные подходы, в которых физическая модель сигнала сочетается с методами машинного обучения. Кроме того, возрастает роль мультимодальных систем, объединяющих визуальные данные с информацией о толщине, составе и подповерхностной структуре покрытия [1, 7, 18].

В последние годы развитие в этой области связано прежде всего с мультисенсорными решениями, методами работы на малых выборках и включением контроля в общий контур управления производством [1, 2, 5]. В результате система контроля всё чаще используется не только для отбраковки изделий, но и для анализа причин дефектов и корректировки технологического процесса.

Заключение

Проектирование системы контроля качества покрытий изделий требует совместного учета физических принципов измерения, особенностей дефектов, методов обработки данных и ограничений производственной среды. Контроль покрытий включает не только поиск видимых поверхностных нарушений, но и оценку толщины, равномерности слоя, его внутренней структуры и признаков, связанных с эксплуатационной надежностью [6, 13–22]. Поэтому для таких задач особенно востребованы неконтактные и неразрушающие методы анализа.

На практике оптическое машинное зрение остаётся основным средством быстрого контроля внешних дефектов. В то же время методы ОКТ, терагерцовой, термографической, вихретоковой, рентгенофлуоресцентной и лазерно-спектроскопической диагностики позволяют получать данные о толщине, структуре и составе покрытия. С учетом этого наиболее рациональным для промышленности является использование комбинированных систем, в которых разные методы дополняют друг друга и работают в общем производственном контуре. В таком подходе контроль служит не только для выявления брака, но и как инструмент повышения устойчивости самого технологического процесса.

APPROACHES TO DESIGNING A QUALITY CONTROL SYSTEM FOR COATING PRODUCTS

IU. S. IAKOVLEV, IU. V. BRUTTAN

Abstract. The article discusses approaches to the design of coating quality control systems for industrial applications. The focus is on non-contact and non-destructive methods, including machine vision, optical, eddy-current, thermographic, and spectral control tools. It is shown that the effectiveness of such systems depends not only on the chosen measurement method, but also on how well the system is integrated into the production process.

Keywords: coating quality control, nondestructive testing, non-contact methods, machine vision, optical coherence tomography, terahertz inspection, active thermography, multi-sensor systems.

Список литературы

1. Tang K., Li Y., Feng K., Liu Y. et al. Coating defect detection in intelligent manufacturing: Advances, challenges, and future trends // *Robotics and Computer-Integrated Manufacturing*. 2025. Vol. 94. Art. 103079.
2. Ma Y., Yin J., Huang F., Li Q. Surface defect inspection of industrial products with object detection deep networks: a systematic review // *Artificial Intelligence Review*. 2024. Vol. 57. Art. 333.
3. Ren Z., Fang F., Yan N., Wu Y. State of the art in defect detection based on machine vision // *International Journal of Precision Engineering and Manufacturing-Green Technology*. 2022. Vol. 9. P. 661-691.
4. Chen Y., Ding Y., Zhao F., Zhang E., Wu Z., Shao L. Surface defect detection methods for industrial products: a review // *Applied Sciences*. 2021. Vol. 11. No. 16. Art. 7657.
5. Bai D. et al. Surface defect detection methods for industrial products with imbalanced samples: a review of progress in the 2020s // *Engineering Applications of Artificial Intelligence*. 2024. Vol. 130. Art. 107697.
6. Giurlani W., Berretti E., Innocenti M., Lavacchi A. Measuring the thickness of metal coatings: a review of the methods // *Coatings*. 2020. Vol. 10. No. 12. Art. 1211.
7. Liu J., Zhang Y., Li H. et al. Non-contact measurement of insulating bearing coating thickness based on multi-sensor combination // *Measurement*. 2023. Vol. 207. Art. 112369.
8. Panjan P., Drnovšek A., Gselman P., Čekada M., Panjan M. Review of growth defects in thin films prepared by PVD techniques // *Coatings*. 2020. Vol. 10. No. 5. Art. 447.
9. Kurc-Lisiecka A., Lisiecki A. Quality assessment and analysis of the causes of coating defects in the powder coating process // *International Journal of Modern Manufacturing Technologies*. 2025. Vol. 17. No. 1. P. 7-16.
10. Zhang J., Cho Y., Kim J. et al. Non-destructive evaluation of coating thickness using water immersion ultrasonic testing // *Coatings*. 2021. Vol. 11. No. 11. Art. 1421.
11. Nguyen T., Hubbard J., Pommersheim J. Surface roughness profile and its effect on coating adhesion and corrosion protection: a review // *Progress in Organic Coatings*. 2020. Vol. 148. Art. 105847.
12. Terek P., Kovačević L., Terek V. et al. Surface roughness and its effect on adhesion and tribological performance of magnetron sputtered nitride coatings // *Coatings*. 2024. Vol. 14. No. 8. Art. 1010.
13. Petersen C.R., Rajagopalan N. et al. Non-destructive subsurface inspection of marine and protective coatings using near- and mid-infrared optical coherence tomography // *Coatings*. 2021. Vol. 11. No. 8. Art. 877.
14. Lawman S., Williams B.M. et al. Scan-less line field optical coherence tomography, with automatic image segmentation, as a measurement tool for automotive coatings // *Applied Sciences*. 2017. Vol. 7. No. 4. Art. 351.
15. Isern L., Waddie A.J., Chalk C. et al. Non-destructive thickness measurement of thermal barrier coatings using terahertz radiation // *Emergent Materials*. 2021. Vol. 4. P. 1547-1557.
16. Lenz M., Mazzon C., Dillmann C. et al. Spectral domain optical coherence tomography for non-destructive testing of protection coatings on metal substrates // *Applied Sciences*. 2017. Vol. 7. No. 4. Art. 364.
17. Hansen R.E., Bæk T. et al. Non-contact paper thickness and quality monitoring based on mid-infrared optical coherence tomography and THz time domain spectroscopy // *Sensors*. 2022. Vol. 22. No. 4. Art. 1549.
18. Zhang X., Wang Z., Wu S. et al. Nondestructive testing of corrosion thickness in coated steel structures with THz-TDS // *Measurement*. 2023. Vol. 217. Art. 113088.
19. Meinschmidt P., Ahlers F., Lang W. Laser scanning thermography for coating thickness inspection // *Engineering Proceedings*. 2021. Vol. 8. No. 1. Art. 17.
20. Meng X., Lu M., Yin W., Bennecer A., Kirk K.J. Evaluation of coating thickness using lift-off insensitivity of eddy current sensor // *Sensors*. 2021. Vol. 21. No. 2. Art. 419.
21. Shanfield Z., Bertin M. An X-ray fluorescence coating gauge for paint thickness // *Advances in X-Ray Analysis*. 2019. Vol. 21. P. 95-102.
22. Basler C. et al. Comparison of laser pulse duration for the spatially resolved measurement of coating thickness with laser-induced breakdown spectroscopy // *Sensors*. 2019. Vol. 19. No. 19. Art. 4133.

ПРОТОТИПИРОВАНИЕ ПОДСИСТЕМЫ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЯ ГОЛОВЫ ОТ ФОНА В СОСТАВЕ СИСТЕМЫ ЦЕНТРАЛИЗОВАННОЙ ВИДЕОАНАЛИТИКИ НА БАЗЕ МАШИННОГО ЗРЕНИЯ БПЛА

М.М. ЛУКАШЕВИЧ¹, А.А. ВОРОНОВ², В.В. ВЕНГЕРЕНКО², М.М. ТАТУР³

1 – Белорусский государственный университет, Республика Беларусь

2 – Объединенный институт проблем информатики НАН Беларуси, Республика Беларусь

3 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 6 апреля 2026

Аннотация. Выполнен обзор вариантов использования прототипа подсистемы сегментации изображения головы от фона в составе централизованной системы видеоаналитики с использованием БПЛА. Предложен алгоритм сегментации, использующий модификацию нейросетевой архитектуры. Показано, что алгоритм работает значительно быстрее и эффективнее, чем лучшие известные алгоритмы сегментации и стандартные нейросетевые архитектуры.

Ключевые слова: семантическая сегментация, видеоаналитика, БПЛА, дистанционное зондирование.

Введение

Мониторинг общественной безопасности предполагает наличие точной и оперативно обновляемой информации об объектах инфраструктуры. В настоящее время системы автоматической видеоаналитики достаточно распространены в мире, лидером по количеству камер видеонаблюдения является Китай, осуществивший теоретически полное покрытие всех крупных городов камерами видеорегистрации с продвинутой системой видеоаналитики, использующей идентификацию по FaceID и быстрый поиск по единой базе данных в 1 млрд. записей. Например, системы Hikvision iVMS-5200 (Китай) и QVR Pro (Тайвань).

Одним из лидеров по внедрению видеонаблюдения в DAS (Domain Awareness System) системы отдельных мегаполисов для обеспечения общественной безопасности являются США. Ниже перечислены некоторые из них: VIDEOMA INTELION, Milesight VMS Enterprise, Rhombus physical security platform, On-Net Surveillance Systems, Cisco MerakiMV Smart Cameras, Arcules Platform, VisionPoint, Verkada Video Surveillance & Management.

Третье место по количеству камер видеорегистрации принадлежит Москве с интеграцией в систему распознавания лиц и идентификации преступников. Основные разработчики необходимых функций – Ntechlab (поиск лица в кадре – детектирование), Visionlabs-Сбер (распознавание) и Tevian (идентификация – сверка с базами данных) или Сфера (при работе на транспорте).

Обработка видеоданных в системах видеонаблюдения позволяет расширить их функциональность. Решаются следующие задачи: детекция пешеходов; лиц людей с последующей идентификацией; автомобилей; номерных знаков автомобилей и номеров вагонов поездов с последующим распознаванием; детекция оставленных без присмотра вещей; слежение за объектом, в том числе и с использованием поворотных камер или камер дрона, с автоматическим управлением; анализ плотности скопления людей и автомобильного трафика, подсчет объектов интереса; целеуказание.

Решение каждой из перечисленных задач требует комплексного подхода, и включает ряд подзадач: предварительная обработка кадров видеопотока; выделение областей движения;

вычисление оптического потока; подавление тряски камеры (antishake); детекция объектов (лиц, людей, автомобилей и т.д.); слежение за объектом (трекинг); идентификация объекта.

Важным параметром алгоритмов обработки видеоданных, используемых в задачах видеонаблюдения, является их вычислительная сложность, так как обработка данных в таких системах ведется в реальном масштабе времени. Поэтому, важной характеристикой алгоритмов является возможность функционирования на недорогом оборудовании встроенных систем.

Для реализации быстрого распознавания лиц в системах видеонаблюдения применяются различные подходы к хранению данных и идентификации: использование единого центра обработки данных (ЦОД) или поиск в локальной базе данных (локальные системы хранения данных, напрямую подключенные к серверам для обработки видеопотоков в реальном времени с использованием ИИ-алгоритмов анализа). При использовании локальных баз данных минимизируются временные потери на сетевые коммуникации и возрастает автономность систем видеоаналитики, что важно в сфере безопасности на транспорте и торговле, в распределенной системе автоматизированного пограничного контроля.

Варианты развертывания

Видеоаналитика со стационарных камер не способна обеспечить отсутствие слепых зон при контроле безопасности, поэтому часто используют дроны, которые могут маневрировать и транслировать видеопоток с камер FHD по радиоканалу с достаточной пропускной способностью. Далее, рассмотрим два варианта развертывания мобильных систем видеонаблюдения с применением беспилотного летательного аппарата (БПЛА) и использованием подсистемы сегментации. Оба они проиллюстрированы на рис.1

Первый вариант для систем с централизованной обработкой. Основные системы видеоаналитики с централизованной обработкой, которые интегрированы с дронами следующие: FaceSix, которая лидирует в идентификации, DJI Analytics универсален по решаемым задачам, TRASSIR, DEDRONE [1-6]. Видеопоток от дрона передается в облачный/вычислительный центр, где выполняется полнофункциональная обработка: стабилизация, детекция лиц (YOLOv8 и более ресурсоемкие нейросетевые архитектуры), сегментация, трекинг, калибровка камер, стереосопоставление, 3-D-реконструкция и идентификация. Такой подход обеспечивает более высокую точность и полноту анализа, но требует стабильной сетевой связи и больших вычислительных ресурсов. Сетевые задержки и пропускная способность остаются критичными ограничениями, особенно при использовании 4G/LTE; 5G пока недоступен в некоторых регионах.

Одним из вариантов использования подсистемы сегментации лица человека на фотографиях и восстановления 3D модели для последующей идентификации по FaceID является описанный выше способ трансляции с дрона видеопотока в высоком разрешении на сервер, обладающий необходимыми вычислительными возможностями для сложной обработки и анализа видео в реальном времени. При анализе основной аппаратно-программной инфраструктуры предлагаемого варианта использования подсистемы сегментации в составе системы централизованной видеоаналитики, следует обратить внимание на используемые кодеки сжатия, варианты организации радиоканала и коммуникационной сети для передачи видеопотока с дрона на сервер. В одном из стандартов, регламентирующих процесс передачи видеопотока с дронов военного назначения определяются уровни сжатия MISM-L9H/L10H для Full HD изображений равные соответственно 80:1 и 20:1. Камеры дронов получают изображение Full HD 1920×1080×(50p) и 1920×1080×(30p). Для сокращения требований к радиоканалу применяются уровни сжатия из стандарта MISM-L10H и 10-разрядное представление для одного пикселя, так для передачи видеопотока Full HD 1920×1080×(30p) требуется пропускная способность 50 Мбит/с, а минимальный канал необходимый для передачи видеопотока, декодируемого при помощи MPEG-2 со степенью сжатия от 5 до 10 составляет от 34 до 100 Мбит/с. Сжатие потока Full HD с коэффициентами до 45:1 и 80:1 соответствует уровню MISM-L9 стандарта. При сжатии на уровне MISM-L9H достаточно радиолинии с пропускной способностью всего 20 Мбит/с для передачи такого видеопотока. В стандарте определены содержание и формат данных, получаемых с радаров обнаружения движущихся целей на фоне земной поверхности (GMTI – Ground Moving Target Indicator). В зависимости от пропускной способности каналов связи описанный в стандарте формат GMTI позволяет передать только

информацию о движущихся целях либо еще и сопутствующие радиолокационные изображения высокого разрешения (рис.1) [5]. Данные транслируются пакетами размером 65535 байт, что позволяет передать информацию о 80 отметках целей и их траекториях, сопровождаемых с высоким разрешением, или о 4300 целях и их трассах, сопровождаемых в обычном режиме. Всего на борту дрона может быть 256 источников данных (основных и вспомогательных) и генерировать они могут данные различных типов: 64 сенсора, формирующие изображение, так же данные от других бортовых датчиков: навигационного оборудования, бортовых систем управления и системы управления видеокамерой на борту дрона. Учитывая мировой опыт на борту дрона эффективно использовать видеокамеры высокой четкости с прогрессивной разверткой и квадратной формой пикселя, например, Full HD 1920×1080×(50p), чтобы не загружать для пересылки видеопотока радиоканал, имеющий ограниченную пропускную способность, целесообразно сжимать видео и использовать коэффициенты сжатия из стандарта соответствующие уровням MISM-L9H/L10H.

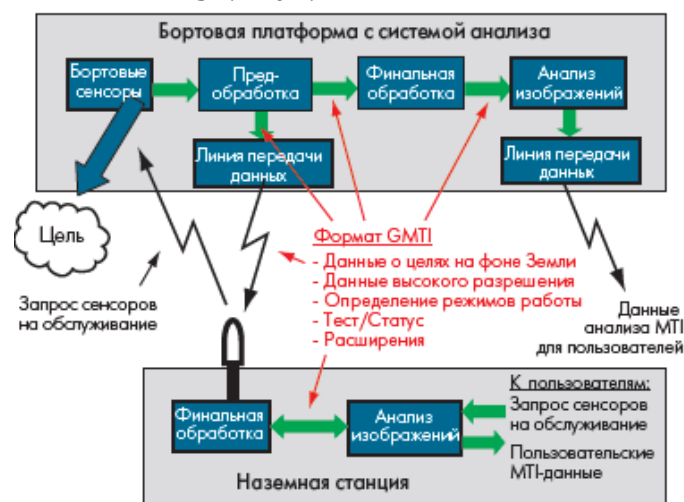


Рис. 1. Архитектура системы передачи данных об объектах интереса с дрона

В условиях ограниченной пропускной способности радиоканала в мобильных системах наблюдения на практике передача видеопотока невозможна без сжатия. Для сжатия отдельных кадров из которых и состоит видеопоток в настоящее время широко используются кодеки стандартов JPEG (Joint Photographic Experts Group), JPEG2000, основанные на алгоритмах энтропийного кодирования Хаффмана и арифметического; стандартов GIF (Graphics Interchange Format), TIFF (Tagged Image File Format), основанные на алгоритме LZW (Lempel-Ziv-Welch); методы и алгоритмы RLE (Run Length Encoding), LZ (Lempel-Ziv), усредненного блочного кодирования, EZW (Embedded Zerotrees of Wavelet transforms), SPIHT (Set Partitioning In Hierarchical Trees), SPECK (Set Partitioned Embedded bloCK) и ряд других [6]. Однако эти кодеки, методы и алгоритмы ориентированы, как правило, на сжатие отдельных изображений, учитывают только кодовую, межэлементную избыточность

Когда имеем дело с видеопотоком используются кодеры, которые могут работать в двух основных режимах: внутрикадрового и межкадрового прогнозирования. Внутрикадровое прогнозирование используется для кодирования опорных кадров. Межкадровое прогнозирование основано на блочной компенсации движения и используется для формирования прогнозных кадров на основе опорных кадров или других прогнозных кадров. Видеодекодер, осуществляет обратное преобразование и сглаживание значений восстановленных пикселей на границах блоков для вычисления ошибки прогнозирования, последующего ее кодирования и передачи. Основные отличия кодеков MPEG-4, H.264, H.265 заключаются в использовании различных преобразований и диапазонах изменения размеров блоков. В видеокодеке MPEG-4 применяется дискретное косинусное преобразование. Видеокодек H.264 использует дискретное вейвлет-преобразование. В видеокодеке H.265 существенно расширен диапазон изменения размеров блоков при блочной компенсации движения. При пиксельной компенсации движения, применяемой в MPEG-2, MPEG-4, H.264, H.265 прогнозный кадр формируется в результате

смещения всех пикселей опорного кадра на одинаковую величину [7]. Это позволяет компенсировать поворот, тангаж и смещение.

Данные передаются по радиолинии пакетами и для каждого пакета вычисляется контрольная сумма CRC. В зонах покрытия сотовой связью используется мобильная сеть. Если нет устойчивого покрытия мобильной связью, то используются специальные радиосети в том числе с ретрансляторами на базе БПЛА обеспечивающие надежную передачу информации с вероятностью ошибки на уровне от 10^{-7} до 10^{-6} , что для FullHD кадра равно одному пикселю на кадр. Для случая возникновения ошибок применяется помехоустойчивое кодирование, которому необходима минимум 30 % избыточность за счет добавления служебной информации при кодировании [8]. Помехоустойчивое кодирование обеспечивает восстановление передаваемой информации при ее искажениях в канале передачи. Если сжатие ориентировано на уменьшение избыточности передаваемой информации, то помехоустойчивое кодирование, наоборот, основано на внесении в передаваемую информацию избыточности, позволяющей приемной стороне обнаружить и исправить ошибки, возникающие в канале передачи. Избыточность в передаваемую информацию, в простейшем случае, вносится за счет добавления проверочных и корректирующих символов, значения которых выбираются в соответствии с алгоритмом помехоустойчивого кодирования. В более сложном случае осуществляется преобразование передаваемой информации. Различные методы помехоустойчивого кодирования обеспечивают различную корректирующую способность и различную избыточность. В общем случае, избыточность увеличивается пропорционально увеличению корректирующей способности. Поэтому, при использовании помехоустойчивого кодирования необходимо предусмотреть соответствующее увеличение коэффициента сжатия при эффективном кодировании для того, чтобы сохранить требуемую пропускную способность канала передачи. Широкое распространение для помехоустойчивого кодирования получили коды Рида-Соломона (для борьбы с ошибками внутри модулей), табличные низкоплотные коды (LDPC-коды, позволяющие ускорить декодирование), BCH-коды (для борьбы со случайными ошибками) и их модификации [9]. Далее приведены примеры технологий, обеспечивающих передачу видеопотока необходимого качества, так для технологии Wi-Fi, например, скорости передачи от 1 Гбит/с до 54 Мбит/с при радиусе зоны уверенного приема 300 м на открытой местности [10]. Для технологии WiMAX скорость передачи 75 Мбит/с при радиусе зоны уверенного приема 25 км [10], а в мобильной версии со скоростью 40 Мбит/с и меньшим радиусом зоны уверенного приема 5 км. При использовании сети мобильной связи 4G(LTE) можно передавать Full HD видеопоток с задержкой, если передача идет по сети 5G передача идет без пауз, но сеть 5G массово развернута только в Китае, США, Южной Корее, ОАЭ и 80 странах Азии, Америки и Европы. Первый вариант использования с развертыванием подсистемы сегментации изображений головы и восстановления 3D модели головы в ЦОД и трансляции качественного видеопотока практически реализуем с большей вероятностью.

Второй вариант развертывания подсистемы сегментации и распознавания лиц во встроенном исполнении на базе Raspberry Pi 5 использующий специализированные облегченные(квантованные) нейронные сети для детектирования (нейронная сеть YOLOv8n квантованная) и идентификации (FaceNet квантованная) соответственно [11]. В данном исполнении видеопоток обрабатывается на борту дрона, что позволяет принимать решения в реальном времени даже в условиях ограниченной связи, соответственно не требуется высокая пропускная способность сети передачи данных. Выполняемые вычисления не должны быть требовательны к ресурсам и адаптированы для запуска на встроенных вычислителях, что делает данный вариант жизнеспособным для мобильного использования, но с ограниченным качеством распознавания, и соответственно такими же ограничениями по применению сценариев 3D моделирования головы человека и идентификации в реальном времени. В тоже время нейронные сети семейства YOLO в последнее время приобрели популярность среди разработчиков и доминируют в решениях для детектирования: улучшилось выделение признаков и разномасштабное обнаружение, появилась возможность применения в автономном вождении, наблюдении и в робототехнике [12]. Чтобы решить задачу идентификации используются нейронные сети с низкой вычислительной сложностью, компактное хранилище данных, и высокую точность распознавания лиц [13] для изображений с изменяющимся качеством при

неблагоприятных погодных условиях. В статье [14] описаны мультимодальные системы, использующие несколько каналов для входных данных различной природы.

Аппаратно-программная база для второго варианта развертывания подсистемы сегментации. Аппаратная основа системы с обработкой видео непосредственно на встроенной системе дрона: Raspberry Pi 5 для выполнения вычислений, камера - источник видеопотока. Основа программной части: квантованная для запуска на Raspberry Pi 5 нейронная сеть YOLOv8n, предварительно обученная на наборе данных COCO для обнаружения лица человека, и также квантованная нейросеть DeepFace (FaceNet) для распознавания лица и поиска в базе данных существующих фото с использованием метрики косинусного подобия. Как можно заметить в предложенном варианте развертывания актуальными для практического использования являются те методы и алгоритмы детекции и идентификации лица человека, которые не требуют больших вычислительных мощностей.

Структура системы видеоаналитики

Структура системы видеоаналитики в которой могут быть использованы предлагаемые алгоритмы сегментации головы человека от фона для последующей реконструкции 3D модели головы человека в реальном времени и развертывания вычислений в ЦОД представлена на рис.2, так как 3D моделирование и поиск нужного ракурса для получения фото вычислительно сложные задачи.

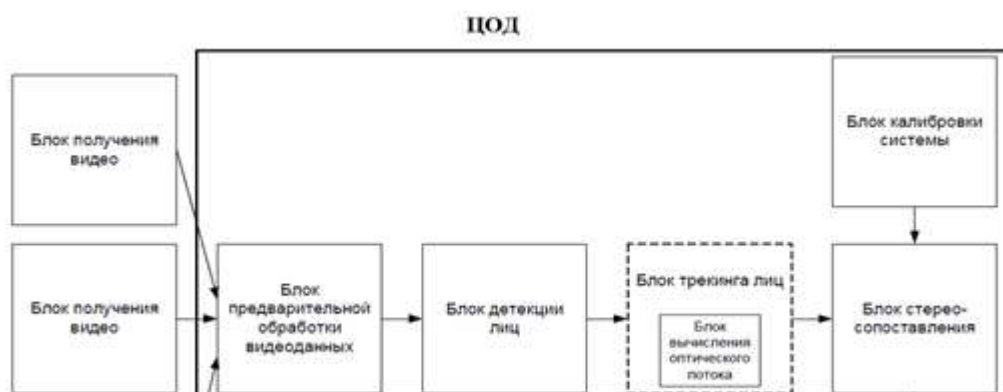


Рис. 2. Архитектура системы видеоаналитики при передаче видеопотока об объектах интереса с дрона в центр обработки данных

Блок получения видео предназначен для получения последовательности кадров видеоизображения. Источником изображения может служить Любой стандартный источник видеопотока (windows video device); видеофайл в формате AVI; графический файл в форматах PNG, JPG, BMP; текстовый файл, содержащий список файлов изображений.

Блок предварительной обработки и стабилизации изображения предназначен для улучшения качества поступающих изображений. В основе данного блока лежат алгоритмы медианной фильтрации для устранения шума. Необходимо устранения возможного дрожания изображения вследствие вибраций камеры с помощью мультимасштабного анализа. Блок детекции лиц и последующей сегментации от фона (выделения оконтуренных объектов) на изображении.

Блок трекинга лиц, вычисление оптического потока предназначен для вычисления областей на изображении, где присутствует визуальное движение, вычисляется усредненная сумма векторов потока по всей области интереса, которая характеризует глобальное перемещение паттерна.

Блок калибровки камер системы предназначен для вычисления фундаментальной матрицы преобразования в соответствии с эпиполярной геометрией. В качестве исходных данных используются точки, указанные пользователем вручную. Вычисленная калибровка сохраняется в файле и может быть многократно использована, но только в том случае, если не изменилось взаимное расположение камер. Калибровка рассчитывается для каждой камеры в системе.

Блок стереосопоставления формирует на выходе стереолица и их 3D модели и выполняет идентификацию лиц непосредственно по их 3D моделям, либо определяет оптимальный ракурс предназначенный для выбора наиболее качественного изображения лица с точки зрения идентификации. Для этого анализируются показатели качества лиц как в многоракурсном домене (стереолицо), так и во временном домене (история наблюдения за лицом).

Методы сегментации лица человека

Методы, применяемые для решения задач сегментации, классификации и идентификация. Сегментация лица человека от фона может использоваться в различных приложениях: например, восстановления 3D модели головы и лица [15], идентификации на основе использования фотографий построенной 3D модели лица и головы с одной стороны и фотографий от системы видеонаблюдения, что и предлагается использовать в экспериментальной системе видеоаналитики, архитектура которой описана ниже и представляет конвейер обработчиков данных.

Методы обнаружения изображения головы на статических кадрах, полученных из видеопотока, можно разделить на следующие:

Эмпирические методы, основанные на знаниях о внешнем виде головы. В таких методах используется набор правил, выполнение которых может привести к решению рассматриваемой задачи. Примерами таких правил могут быть для случая фронтального изображения головы, следующие: лицо имеет симметричную форму, центральная часть лица имеет практически не изменяющуюся текстуру. Как правило, они заключаются в установлении простых закономерностей, например, для фронтального изображения головы: лицо человека обычно симметрично, глаза, нос и рот, как правило отличаются по яркости и т.д. В качестве примеров можно назвать иерархические и проекционные методы [15].

Признаковые методы основаны на поиске отличительных особенностей лица на изображении и сопоставлении найденных признаков для восстановления позиции лица в целом. В качестве отличительных признаков могут рассматриваться контуры лица текстуры, цвет кожи, форма.

Методы моделирования предполагают использование некоторой модели лица, которое сопоставляется с представленным изображением. Модели могут состоять из контуров, характерных точек. В литературе встречаются подходы с жесткими и деформируемыми моделями.

Методы внешнего вида опираются на инструментарий распознавания образов. При этом классификатор настраивается на решение двух классов задачи распознавания. Общее у методов, основанных на внешнем виде лица, состоит в алгоритме получения тестовых областей на изображении. Так как изображение лица может находиться в любом месте кадра и иметь любой размер, то для рассмотрения всех возможных вариантов необходимо перебрать все области на изображении при всех подходящих масштабах.

Каждая такая область рассматривается как тестовый образ и подается на вход классификатора. В качестве методов классификации могут использоваться: факторный анализ (FA – Factor Analysis); линейный дискриминантный анализ (LDA – Linear Discriminant Analysis); скрытые Марковские модели (HMM - Hidden Markov Models); активные модели внешнего вида (AAM – Active Appearance Models); метод опорных векторов (SVM – Support Vector Machines); разреженная сеть просеивающих элементов (SNoW - Sparse Network of Winnows); собственные лица (Eigenfaces); искусственные нейронные сети (NN – Neural Networks). Кроме того, нейросетевые методы могут использоваться для определения ориентации изображения лица, а также для каскадирования слабых классификаторов [16].

Последние два метода рассматриваются сегодня как лучшие с точки зрения высокого процента верного обнаружения лиц и низкого процента ложных детекций. Нейросети давно с успехом применяются для решения поиска локальных экстремумов является базовой операцией для множества задач обработки изображений.

Нейросетевые методы позволяют добиться высокого процента верного обнаружения лиц и низкого процента ложных детекций. Нейросети давно с успехом применяются для решения многих задач в области машинного зрения. Недостатком нейросетевых методов является их

вычислительная сложность. Они не применимы в чистом виде для решения задач детекции лиц в реальном времени. Для решения этой проблемы предпринимаются различные шаги. Например, в [16] для обеспечения работы детектора лиц на базе сверточной нейронной сети в реальном времени используется специализированный нейросетевой процессор.

Алгоритм сегментации изображений головы человека от фона

Алгоритмы удаления фона на изображениях головы человека используют такие архитектуры глубоких сверточных нейронных сетей, как U-Net, DeepLab-V3+ и Mask R-CNN, которые показывают высокое качество сегментации, а для работы на мобильных устройствах и в реальном времени предложены оптимизированные архитектуры: LiteUNet, MODNet и PortraitNet. Они сочетают в себе приемлемое качество сегментации и высокую скорость обработки, что делает их подходящими для внедрения в мобильные потребительские устройства. Одним из наиболее перспективных направлений является использование Transformer-архитектур, таких как Segmenter, Trans-UNet, Swin-Unet, которые особенно эффективны в обработке информации [17–23]. В ходе реализации проекта разработаны и экспериментально исследованы алгоритмы сегментации при помощи нейросетевых архитектур UNet, PortraitNet, Segmenter, SAM. Модели U-Net, PortraitNet, Segmenter, SAM были обучены на наборе данных, представленном исходными изображениями головы человека в разных ракурсах и эталонными масками сегментации. Модель SAM использовалась без дообучения с предварительно расстановкой опорных точек для сегментации. Результаты тестирования показали, что наилучшую точность демонстрирует модель Segmenter ($IoU = 0,9739$, $Dice(F1) = 0,9867$, $Accuracy = 0,9973$), табл 1.

При вычислениях метрик точности IoU (Intersection over Union) или $Jaccard Index$ – отношение пересечения (предсказанной и истинной масок) к их объединению и $Accuracy$ (точность) – доля правильно классифицированных пикселей и $Dice$ коэффициента ($F1$ метрика) приняты следующие обозначения: TP (True Positive) – пиксели, правильно предсказанные как объект; TN (True Negative) – пиксели, правильно предсказанные как фон; FP (False Positive) – пиксели, ошибочно предсказанные как объект (ложные срабатывания).

$Dice$ коэффициент вычисляется по формуле (1), IoU и $Accuracy$ по формулам (2, 3)

$$Dice = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN}, \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

Приведем описание алгоритма для нейросетевой архитектуры Segmenter.

1. Разбиение изображения на патчи. Входное изображение разбивается на фиксированные блоки (патчи) одинакового размера, например, 16×16 пикселей. Каждый патч преобразуется в вектор признаков с помощью линейного слоя.

2. Добавление позиционных эмбеддингов. Каждому визуальному токenu (вектору патча) добавляется позиционный эмбеддинг, который указывает на его местоположение в исходном изображении. Это позволяет модели учитывать пространственную структуру изображения.

3. Обработка через Transformer-encoder. Последовательность токенов подается в Transformer-encoder, где каждый токен взаимодействует со всеми остальными через механизм внимания (self-attention). На этом этапе модель выделяет глобальные признаки и устанавливает связи между различными частями изображения.

4. Инициализация масочных токенов. Decoder использует набор обучаемых масочных токенов (learnable mask tokens), каждый из которых отвечает за формирование определённой семантической маски (например, для каждого класса объекта).

5. Восстановление масок через кросс-внимание. Масочные токены взаимодействуют с закодированными признаками из encoder'a через кросс-внимание, что позволяет decoder'у формировать детализированные маски, соответствующие реальной структуре изображения.

6. Генерация финальной карты сегментации. На выходе decoder'a получается набор масок, каждая из которых соответствует определённому объекту или классу. Эти маски объединяются в финальную карту сегментации, совпадающую по размерам с исходным изображением.

Табл. 1. Результат тестирования моделей нейронных сетей сегментации

Архитектура	IoU	Dice	Accuracy
U-net	0,9307	0,9641	0,9927
PortraitNet	0,9159	0,9560	0,9914
Segmenter	0,9739	0,9867	0,9973
SAM	0,6634	0,7931	0,9597

Заключение

Доклад посвящен рассмотрению возможных вариантов использования(развертывания) подсистемы сегментации изображения головы человека от фона на фотографиях и выбору перспективных алгоритмов для программной реализации в подсистеме. Описан разработанный алгоритм подготовки данных из видеопотока (удаления фона на изображениях головы человека), являющийся важной частью решения задачи реконструкции 3D модели головы человека по видеопотоку, выполнено экспериментальное исследование алгоритмов удаления фона. Выполнен анализ существующих подходов к семантической сегментации изображений головы человека на фотографиях для последующей 3D реконструкции модели головы человека. Установлено, что для условий мобильной съемки с борта БПЛА наиболее перспективным является извлечение ключевых кадров, их масштабирование с сохранением пропорций и оценка резкости с использованием оператора Лапласа, что позволяет эффективно отбирать кадры для дальнейшей сегментации. Для решения задачи удаления фона (бинарной сегментации) протестированы современные архитектуры нейронных сетей: U-Net, PortraitNet, Segmenter, SAM. Модели U-Net, PortraitNet и Segmenter, которые были обучены на наборах данных, представленных исходными изображения головы человека в разных ракурсах и эталонными масками сегментации. Модель SAM использовалась без дообучения с предварительной расстановкой опорных точек для сегментации. Результаты тестирования показали, что наилучшую точность демонстрирует модель Segmenter, а вариант использования или развертывания подсистемы сегментации с практической точки зрения наиболее перспективен с использованием значительных вычислительных мощностей центра обработки данных.

Работа выполнена в рамках договора с БРФФИ Ф25ТУРГ-001 от 05 марта 2025 г., тема "Приложение для генерации компьютерной 3D модели головы человека и ее сравнение с изображениями, полученными с камеры".

PROTOTYPING A HEAD-BACKGROUND SEGMENTATION SUBSYSTEM AS PART OF A CENTRALIZED VIDEO ANALYTICS SYSTEM BASED ON UAV MACHINE VISION

M.M. LUKASHEVICH, A.A. VORONOV, V.V. VENGERENKO, M.M. TATUR

Abstract. The article is devoted to not only processing aerial photographs of human face for the task of safety monitoring, but also to the review of possible use cases for a prototype head-background segmentation subsystem within a centralized UAV-based video analytics system and development of algorithms for processing and segmentation of photoimages of human face for the task of safety monitoring. Segmentation algorithm using a modified neural network architecture is proposed. It is shown to be significantly faster and more efficient than the best known segmentation algorithms and state of art neural network architectures.

Keywords: semantic segmentation, video analytics, UAVs, incident detection, classification.

Список литературы

1. Ye Z., Lan C. [et al.] // Journal of Visual Communication and Image Representation. 2023. Vol. 95, P. 103903–103910.
2. Yin N., Liu C. [et al.] // IEEE Transactions on Multimedia. 2024. Vol. 26. P. 7812–7822.
3. Zhou G., Xu Q. [et al.] // IEEE Transactions on Geoscience and Remote Sensing. 2025. Vol. 63. P. 1–13.
4. Wang Y., Yang Y. [et al.] // Proc. of European Conference on Computer Vision. 2020. P. 651–664.
5. Слюсар В. // Электроника НТБ Часть 1. 2009. № 5. С. 68–73.
6. Lazzaroni F., Leonardi R. [et al.] // IEEE Signal Processing Letters. 2003. Vol. 10. P. 28–31.
7. Deniz O., Bueno G. [et al.] // Pattern Recognition. Vol. 44. P. 2887–2901.
8. Цветков В. // Доклады БГУИР. 2019. № 3(121). С. 25–36.
9. Sullivan G., Ohm J. [et al.] // IEEE Transactions on Circuits and Systems for Video Technology. 2012. Vol. 22. P. 1649–1668.
10. Вишневский В., Портной С. [и др.] Энциклопедия WiMax. Путь 4G. М., 2009.
11. Yang Y., He H. [et al.] // Algorithms. 2025. Vol. 18. P. 453–460.
12. Nagpal S., Etnubbarhi N. [et al.] // Proc. of Innovations in Intelligent Systems and Applications Conference (ASYU). 2024. P. 1–6.
13. Ren Z., Liu X. [et al.] // Journal of Imaging. 2025. Vol. 11. P. 24–30.
14. Wu W., Chen X. [et al.] // Proc. of International Joint Conference on Neural Networks (IJCNN). IEEE. 2024. P. 1–8.
15. Вежнев В. // Математические методы распознавания образов (ММРО-10): 10-я Всероссийская конференция. 2001. С. 179–180.
16. McRaven J., Scheutz M. [et al.] // IEEE international workshop Cellular Neural Networks and their Applications (CNNA 2004). 2004. P. 196–201.
17. Wu C. // Proc. of the International Conference on 3D Vision (3DV). 2013. P. 127–134.
18. Yao Y., Luo Z. [et al.] // Proc. of the European Conference on Computer Vision (ECCV). 2018. P. 767–783.
19. Gu X., Fan Z. [et al.] // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 3265–3274.
20. Choy C., Xu D. [et al.] // Proc. of the European Conference on Computer Vision (ECCV). 2016. P. 628–644.
21. Qi C., Yi L. [et al.] // Proc. of conf. Advances in Neural Information Processing Systems (NIPS). 2017. P. 5099–5108.
22. Mildenhall B., Srinivasan P. [et al.] // Proc. of the European Conference on Computer Vision (ECCV). 2020. P. 405–421.
23. Saito S., Huang Z. [et al.] // Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 2304–2314.

УДК 004.65:61

МНОГОУРОВНЕВАЯ АРХИТЕКТУРА БЛОКЧЕЙНА ДЛЯ ХРАНЕНИЯ МЕДИЦИНСКИХ ДАННЫХ

В.А. ВИШНЯКОВ, М.О. КАЦКО

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь**Поступила в редакцию 6 апреля 2026*

Аннотация. Проанализированы международные разработки в области блокчейн-систем для здравоохранения. Предложена многоуровневая архитектура подсистемы блокчейн для хранения медицинских данных, объединяющая распределенное хранилище IPFS, гибридную блокчейн-инфраструктуру, систему атрибутивного управления доступом и клиентские приложения. Представлена структурная схема системы с описанием функциональных модулей. Разработан алгоритм работы системы, включающий регистрацию пользователей, загрузки данных, предоставления доступа и верификации транзакций.

Ключевые слова: блокчейн, медицинские данные, IPFS, архитектура, децентрализованное хранение, смарт-контракты, конфиденциальность, безопасность данных.

Введение

Цифровая трансформация здравоохранения привела к экспоненциальному росту объемов медицинских данных. Электронные медицинские карты (ЭМК), данные диагностического оборудования, результаты лабораторных исследований, информация с носимых устройств и телемедицинских платформ формируют сложные информационные массивы, требующие надежного хранения и защищенного обмена. Традиционные централизованные системы хранения медицинских данных сталкиваются с рядом фундаментальных проблем: рисками утечек и несанкционированного доступа, высокими затратами на обеспечение безопасности, сложностью интероперабельности между различными медицинскими учреждениями, а также отсутствием у пациентов контроля над собственными данными [1].

Технология блокчейн предлагает принципиально иной подход к организации хранения медицинских данных. Ее ключевые характеристики – децентрализация, неизменяемость записей, криптографическая защита и прозрачность транзакций, что делают ее идеальной основой для систем управления медицинской информацией [2]. Однако прямое хранение медицинских записей в блокчейне (on-chain storage) сталкивается с ограничениями по объему данных, высокой стоимостью транзакций и проблемами масштабируемости. Это обуславливает необходимость разработки гибридных архитектур, сочетающих блокчейн для обеспечения неизменяемости и контроля доступа с распределенными файловыми системами (IPFS) для хранения самих медицинских данных [3]. Целью настоящей работы является разработка многоуровневой архитектуры подсистемы блокчейн для хранения медицинских данных, объединяющей передовые международные подходы и обеспечивающей баланс между безопасностью, масштабируемостью и удобством использования для пациентов и медицинских учреждений.

Международные разработки

Проект STORChain (IEEE Transactions on Services Computing, 2026) представляет собой оптимизированную по хранению блокчейн-структуру, разработанную для медицинских данных. Ключевой инновацией является Clustered Merkle Patricia Tree (C-MPT) – кластеризованное

дерево Меркла-Патриции, которое агрегирует транзакции схожих типов для максимальной эффективности хранения при сохранении возможности доказательства включения (Proof of Inclusion) [4]. STORChain включает: C-MPT структуру: объединение похожих типов транзакций в кластеры для уменьшения накладных расходов; Selective Transaction Pruning Strategy (STPS): стратегию выборочной обрезки транзакций для приоритизации и удаления устаревших исторических данных Делегированное доказательство доли (DPoS): алгоритм консенсуса с вероятностным механизмом выборов для повышения справедливости и инклюзивной узлов.

Трехуровневая архитектура TLSA (Wiley, 2025). Исследователи K. Maithili и S. Amutha предложили Tri-Layered Sharding Architecture (TLSA) – иерархическую модель, организующую сеть в три шардинговых уровня: Transaction Layer (уровень транзакций), Data Layer (уровень данных) и Location Layer (уровень расположения) [5].

Система CareChain, использующая два блокчейна (публичный для пациентов и приватный для медицинских работников) в сочетании с распределенным хранилищем IPFS [6]. Ключевые особенности CareChain: двойная блокчейн-структура: разделение данных пациентов и провайдеров для повышения конфиденциальности; IPFS распределенное хранение: снижение задержки транзакций и накладных расходов на хранение; ECDSA (Elliptic Curve Digital Signature Algorithm): подпись транзакций для обеспечения подлинности; устройство-прокси: мониторинг низкопроизводительных IoT-устройств для выявления уязвимостей. Результаты тестирования показывают повышение устойчивости системы по сравнению с существующими моделями здравоохранения, а также улучшение требований к хранилищу, энергоэффективности, скорости транзакций, конфиденциальности и безопасности [6].

Minima, Siemens, ARM: блокчейн-на-кристалле (2025-2026), направлен на создание первого в мире промышленного микрочипа, способного выполнять полноценный узел блокчейна [7]. Результаты проекта: 100-кратное ускорение хеширования за счет SHA-3 аппаратного ускорителя; пятикратное снижение использования памяти после рефакторинга C++ кода; Возможность работы устройства как самостоятельного полного узла без обращения к облачной инфраструктуре [7].

Индийский институт технологий (ИИТ Madras) разработал BlockTrack – блокчейн-систему для мобильных приложений, которая в настоящее время проходит полевые испытания в университетской больнице [8]. Характеристики BlockTrack: децентрализованный обмен медицинскими данными для мобильных приложений; Уникальные идентификационные коды для пользователей с низкой вероятностью дублирования.

Проект Emtruth, поддержанный Национальным научным фондом США (NSF) на сумму почти 1 млн долларов, разрабатывает платформу для интеграции медицинских данных с использованием блокчейна и искусственного интеллекта [2]. Ключевые цели проекта: безопасное хранение данных любого типа из любых источников в неизменяемых блокчейнах; сохранение контроля данных за владельцем с возможностью безопасного предоставления доступа; трансформация и нормализация данных в гранулярные блокчейны для индивидуального или агрегированного моделирования.

В Университете ИТМО предложена трехкомпонентная архитектура, включающая блокчейн для хранения зашифрованных медицинских записей, централизованный сервер для справочной информации и клиентское приложение для управления доступом [3]. Особое внимание уделяется оптимизации EVM-блокчейна через адаптацию алгоритма консенсуса, снижение времени блока и устранение транзакционных комиссий.

Структура системы

На основе анализа международных разработок разработана многоуровневая архитектура подсистемы блокчейн для хранения медицинских данных, которая объединяет пять функциональных модулей, каждый из которых решает специфические задачи обеспечения безопасности, масштабируемости и доступности данных. Архитектура представлена на рис.1.

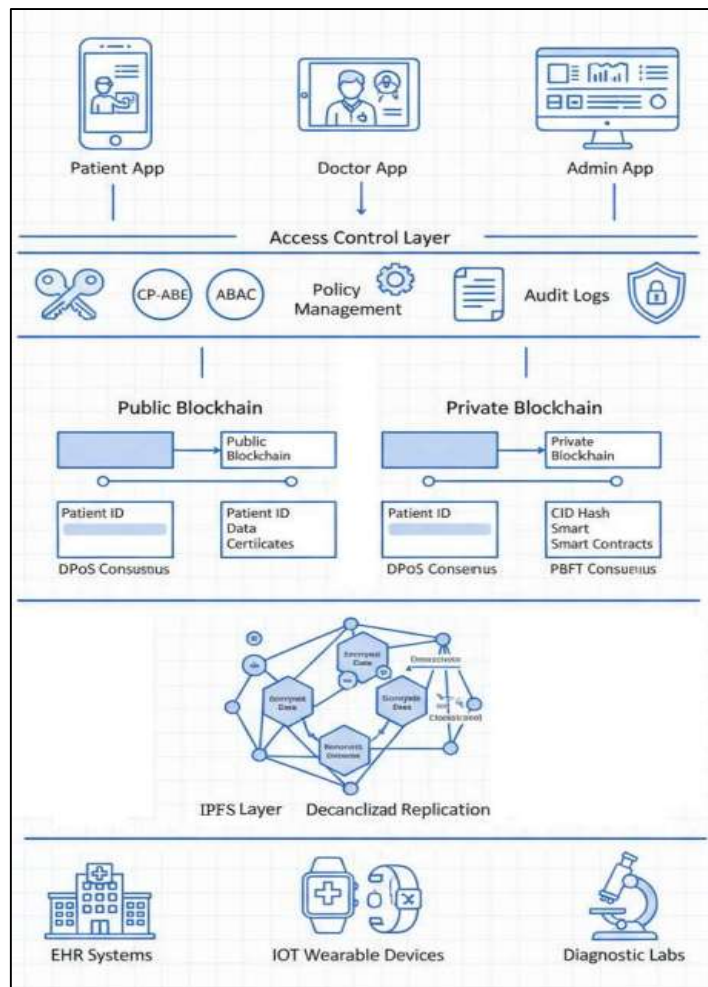


Рис. 1. Многоуровневая архитектура подсистемы блокчейн

Рассмотрим их назначение каждого уровня.

1. Источники данных (Data Sources Layer): медицинские учреждения, генерирующие электронные медицинские записи и диагностические изображения; IoT-устройства и носимые сенсоры, собирающие физиологические показатели в реальном времени; диагностические лаборатории, предоставляющие результаты анализов и исследований;

2. Распределенное хранилище IPFS (IPFS Layer): децентрализованное хранилище, адресуемое по содержимому; хранение зашифрованных медицинских записей, результатов исследований и данных IoT-устройств; генерация уникальных Content Identifiers для ссылок на данные; кэширование и репликация для повышения доступности и снижения задержек [6].

3. Блокчейн-инфраструктура (Blockchain Layer): гибридная архитектура, объединяющая публичный и приватный блокчейны [4]; публичный блокчейн: идентификаторы пациентов, сертификаты учреждений, информация о роуминге; приватный блокчейн: хеши записей (CID), транзакции доступа, аудиторские журналы, смарт-контракты; алгоритмы консенсуса: DPoS для масштабируемости, PBFT для приватных сетей [4].

4. Управление доступом (Access Control Layer): система атрибутивного управления доступом на основе CP-ABE и ABAC [7]; генерация и управление криптографическими ключами; определение политик доступа на основе атрибутов пациентов, врачей и учреждений; журналирование всех операций доступа для аудита;

5. Приложения (Application Layer): пациентское приложение: просмотр записей, управление доступом, предоставление согласий; врачебное приложение: доступ к записям пациентов, добавление заключений, назначение лечения; административное приложение: управление политиками, аудит, регистрация учреждений.

Описание модулей

Шифрования и защиты данных обеспечивает криптографическую защиту медицинских данных на всех этапах обработки. Основные компоненты: для хранения в IPFS шифрование на основе CP-ABE или 128-битное AES с последовательностью ДНК-кодирования [9]; для передачи используется цифровой конверт RSA 2048 для безопасной передачи симметричных ключей; цифровая подпись RSA 1024 для обеспечения подлинности данных; ECDSA: генерация пар публичных и приватных ключей для устройств, пациентов и провайдеров; подпись транзакций для обеспечения подлинности и защиты от подделки; верификация подписей при получении данных [4]; устройство-прокси (Device Proxy): мониторинг низкопроизводительных IoT-устройств для выявления уязвимостей; непрерывная проверка типа устройства, производителя, версии прошивки, безопасности; обнаружение подозрительного поведения через тестирование на проникновение и поведенческий анализ [6].

Хранения IPFS реализует распределенное хранение медицинских данных с адресацией по содержимому. Основные функции: Content-Addressable Storage: данные идентифицируются по их криптографическому хешу (CID), что обеспечивает неизменяемость и возможность верификации [6]; Distributed Hash Table (DHT): распределенная хеш-таблица для хранения соответствия между CID и сетевыми адресами узлов, содержащих данные; кэширование и репликация: кэширование часто запрашиваемых данных на соседних узлах для снижения задержек и уменьшения нагрузки на сеть; Content Identifier (CID) Management: генерация и хранение CID для загружаемых данных; CID сохраняются в блокчейне, а данные – в IPFS [1].

Смарт-контракты реализуют автоматизированную логику управления доступом и обработки транзакций. Основные функции: управление доступом: смарт-контракты проверяют атрибуты пользователя и политики доступа перед предоставлением CID [8]; обработка согласий: автоматическое отслеживание предоставленных пациентом согласий на доступ к данным, их сроков действия и отзыва; аудит и журналирование: запись всех операций доступа в неизменяемый журнал блокчейна для последующего аудита; управление идентификаторами: создание и верификация уникальных идентификаторов пациентов и медицинских учреждений [8].

Управления идентификацией обеспечивает уникальную идентификацию пациентов и медицинских учреждений в распределенной системе. Основные функции: генерация уникальных идентификаторов: алгоритмы, обеспечивающие уникальность ID с низкой вероятностью дублирования [8]; связывание идентификаторов: возможность связывания идентификаторов пациента из разных систем для создания полной медицинской карты [2]; кросс-учрежденческая идентификация: обеспечение интероперабельности между различными медицинскими организациями; управление сертификатами: хранение публичных ключей и сертификатов учреждений в публичном блокчейне.

Аудит и мониторинг обеспечивает прозрачность и возможность проверки всех операций с медицинскими данными. Основные функции: *неизменяемый журнал доступа*: все операции доступа фиксируются в блокчейне и не могут быть изменены или удалены [1]; *анализ аномалий*: обнаружение подозрительных паттернов доступа для выявления потенциальных нарушений безопасности [6]; *генерация отчетов* о доступе для пациентов, врачей; *уведомления*: оповещение пациентов о каждом случае доступа к их данным.

Алгоритм работы системы

Разработаны алгоритмы взаимодействия участников в рамках многоуровневой блокчейн архитектуры, обеспечивающие неизменность записей и управляемый доступ.

Алгоритм регистрации пользователя состоит из следующих шагов:

1. Пациент или врач загружает клиентское приложение и заполняет регистрационную форму.
2. Приложение генерирует пару криптографических ключей (приватный и публичный) с использованием ECDSA [6].
3. Приложение отправляет регистрационную информацию и публичный ключ в систему. Данные верифицируются, и создается уникальный идентификатор [8].

4. Идентификатор и публичный ключ записываются в публичный блокчейн. Создается запись в приватном блокчейне с начальными настройками доступа.

5. Пользователю предоставляется доступ к приложению с возможностью управления своими данными.

Алгоритм загрузки медицинских данных включает в себя этапы:

1. Загрузка данных (ЭМК, результаты анализов, изображения) медицинским учреждением или врачом через приложение.

2. Шифрование данных с использованием гибридной схемы (AES или CP-ABE) в соответствии с атрибутами доступа [9].

3. Передача зашифрованных данных в IPFS-сеть. IPFS вычисляет CID на основе содержимого и распределяет данные по узлам [8].

4. Возврат CID (Content Identifier) – уникальную ссылку на данные в IPFS.

5. Запись CID и метаданных в приватный блокчейн через смарт-контракт [8].

6. Добавление записи о факте создания новой записи (без раскрытия содержимого) в публичный блокчейн.

Алгоритм предоставления доступа к данным описывается следующим образом:

1. Пациент через приложение выбирает медицинского специалиста или учреждение для предоставления доступа.

2. Пациент определяет параметры доступа, объем данных, срок действия, цель использования [2].

3. Приложение формирует политику доступа в формате атрибутов (ABAC) или шифрует данные с политикой CP-ABE [12].

4. Смарт-контракт в приватном блокчейне регистрирует политику доступа и связывает ее с идентификаторами пациента и получателя.

5. Получатель уведомляется о предоставленном доступе через приложение. При необходимости получатель проходит дополнительную аутентификацию.

6. Все действия по предоставлению доступа фиксируются в аудиторском журнале блокчейна.

Алгоритм доступа к данным:

1. Врач или медицинское учреждение через приложение запрашивает доступ к данным пациента.

2. Приложение направляет запрос к смарт-контракту для проверки наличия действующего согласия.

3. Смарт-контракт проверяет политики доступа, соответствие атрибутов запрашивающего, срок действия согласия, объем разрешенных данных [9].

4. При успешной проверке смарт-контракт возвращает cid запрошенных данных.

5. Приложение запрашивает данные из IPFS по полученному CID. IPFS-сеть извлекает зашифрованные данные [4].

6. Данные расшифровываются с использованием ключей, соответствующих политике доступа, и отображаются врачу.

7. Факт доступа регистрируется в блокчейне с указанием времени, запрашивающего и объема данных.

Алгоритм отзыва доступа предусматривает следующие ключевые шаги:

1. Пациент через приложение инициирует отзыв предоставленного доступа.

2. Приложение формирует транзакцию отзыва с указанием идентификатора.

3. Смарт-контракт деактивирует политику доступа, делая ее недействительной.

4. Смарт-контракт обновляет аудиторский журнал, фиксируя факт отзыва.

5. Получатель уведомляется об отзыве доступа. Последующие попытки доступа будут отклонены.

6. Может быть выполнено повторное шифрование данных с новыми политиками доступа.

Особое внимание в предложенном алгоритме работы системы уделено обеспечению конфиденциальности, неизменяемости записей и ограничениям в предоставлении доступа.

Заключение

Международные разработки демонстрируют высокую активность и значительные достижения. Предложенная многоуровневая архитектура объединяет лучшие мировые практики в единое решение. Пятиуровневая структура (источники данных, IPFS, блокчейн, управление доступом, приложения) обеспечивает разделение ответственности, масштабируемость и безопасность. Использование гибридной блокчейн-архитектуры позволяет сохранить прозрачность идентификации при обеспечении конфиденциальности медицинских записей.

Ключевыми компонентами системы являются. модуль шифрования на основе ECDSA и CP-ABE, обеспечивающий защиту данных; модуль IPFS для распределенного хранения с адресацией по содержимому; модуль смарт-контрактов для автоматизации управления доступом; модуль управления идентификацией для уникальной идентификации пациентов; модуль аудита для обеспечения прозрачности операций.

MULTILEVEL BLOCKCHAIN ARCHITECTURE FOR MEDICAL DATA STORING

U.A VISHNIAKOU, M.A. KATSKO

Abstract. International developments in the field of blockchain systems for healthcare are analyzed,. A multi-level architecture of the blockchain subsystem for storing medical data is proposed, combining IPFS distributed storage, hybrid blockchain infrastructure, attribute access control system and client applications. A structural diagram of the system with a description of the functional modules is presented. The system's algorithm has been developed, which includes the stages of user registration, data upload, access, and transaction verification. Keywords. blockchain, medical data, IPFS, architecture, decentralized storage, smart contracts, privacy, data security.)

Keywords. blockchain, medical data, IPFS, architecture, decentralized storage, smart contracts, privacy, data security.

Список литературы

1. Adapa, C. S. R. Enhancing Healthcare Data Integrity Through Blockchain-Based Master Data Management. An Architectural Framework / C. S. R. Adapa // Journal of Information Systems Engineering and Management. – 2026. – Vol. 11.
2. Emtruth, Inc. SBIR Phase II. A Platform for Health Care Data Integration Using Blockchain and Artificial Intelligence / NSF Award 2304102. – National Science Foundation, 2023. – URL. <https://www.sbir.gov/awards/2304102> (дата обращения. 16.03.2026).
3. Лаврова, А. К., Максимова, Т. Г. Блокчейн для медицины. новая модель хранения и управления медицинскими данными / А. К. Лаврова, Т. Г. Максимова // Сборник тезисов докладов конгресса молодых ученых. – СПб . Университет ИТМО, 2025. – URL. <https://kmu.itmo.ru/digests/article/14096> (дата обращения. 16.03.2026).
4. STORChain. A Clustered-MPT-Based Blockchain for Data Service and Efficient Storage in Healthcare / IEEE Transactions on Services Computing. – 2026.
5. Maithili, K., Amutha, S. Optimizing Blockchain Scalability for Secure Patient Health Records With Tri-Layered Sharding Architecture (TLISA) / K. Maithili, S. Amutha // Transactions on Emerging Telecommunications Technologies. – 2025.
6. CareChain. Secure and scalable dual blockchain and IPFS driven IoT ecosystem for next gen healthcare systems / Scientific Reports. – 2025.
7. Minima Achieves Major Breakthrough. Blockchain-on-Chip Is Here / Minima Global. – 15 December 2025. – URL. <https://minima.global/ru/post/minima-achieves-major-breakthrough-blockchain-on-chip-is-here> (дата обращения. 16.03.2026).
8. IIT Madras. IIT Madras Develops Blockchain-based Healthcare Information Systems for Mobile Apps / Principal Scientific Adviser, Government of India. – 2026. – URL. <https://www.psa.gov.in/article/iit-madras-develops-blockchain-based-healthcare-information-systems-mobile-apps/2744> (дата обращения. 16.03.2026).
9. Sanobar, A., Anwar, S. A secure and privacy preserving model for healthcare applications based on blockchain-layered architecture / A. Sanobar, S. Anwar // International Journal of Computers and Applications. – 2024. – Vol. 46. – No. 12. – P. 1206-1218.

АЛГОРИТМЫ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ CLC И OPCOSA

Т. А. ТУМАНОВ, Л. ВЕЙ, А. Д. ВАСЮТИЧ, Ц. МА, С. ЖЭНЬ, А. В. КУРИЛОВИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 6 апреля 2026

Аннотация. Проведено сравнение алгоритмов двумерного помехоустойчивого кодирования и декодирования информации, представленной в виде матрицы 4×4 : Collum-Line Codes (CLC) и Optimized Product Codes for Space Applications (OPCoSA). Алгоритмы реализованы в программе моделирования цифровых логических схем Logisim и среде проектирования Vivado для программируемых логических интегральных схем (ПЛИС) Xilinx. Получены количественные оценки занимаемых кодеками ресурсов и потребляемой мощности ПЛИС.

Ключевые слова: коды с коррекцией ошибок, CLC, OPCoSA, моделирование, комбинации многократных ошибок.

Введение

Так как в процессе передачи по каналам связи информация может исказиться, из-за действия внешних факторов, приводящих к одиночным или множественным ошибкам, разработаны и широко применяются методы помехоустойчивого кодирования и декодирования, позволяющие обнаружить и устранить несоответствия в принятом сообщении.

В случае передачи информации в космическом пространстве фактором, влияющим на сигнал, могут быть высокоэнергетические радиационные частицы окружающего пространства. Во многих случаях, действие этого фактора, приводит к нежелательным последствиям, которые уменьшаются при использовании кодеков с коррекцией ошибок [1, 2]. На сегодняшний день кодеки с возможностью корректирования только одиночных ошибок устарели, так как с развитием технологий передачи информации они перестали удовлетворять потребностям. В связи с этим разработаны различные варианты алгоритмов, способные восстановить информацию в случае возникновения многократных ошибок [3]. Представителями этой группы являются кодеки Collum-Line Codes (CLC) и OPCoSA, построенные на базе расширенного кода Хэмминга с проверкой на четность. Они обладают высокой корректирующей способностью при сравнительно небольшом потреблении вычислительных ресурсов приемно-передающих устройств [4–6].

Коррекция ошибок в CLC и OPCoSA

Кодек CLC (16, 40) представляет собой двунаправленную матрицу кодирования, которая формирует контрольные биты по столбцам и биты проверки четности по строкам. Для CLC 40-битное кодовое слово, состоит информационных бит в виде матрицы 4×4 , которая дополняется проверочными разрядами (8, 4) расширенного кода Хэмминга по строкам и для каждого столбца дополняется битом проверки четности [4]. Итоговая структура позволяет корректировать одиночные ошибки с помощью алгоритма Хэмминга и множественные ошибки с помощью битов проверки четности. Кроме того, есть и другие разновидности кода CLC, такие как CLC (16, 39) и CLC (16, 54) [5]. Отличие CLC (16, 39) от CLC (16, 40) заключается в том, что данные 16 информационных битов записываются в 2 строки по 8 бит, к которым применяется расширенный код Хэмминга и добавляются биты проверки четности в столбец и строку. Для CLC (16, 54) применяется следующий алгоритм: 16 бит информации распределяются

в 8 строк по 2 бита, далее аналогично применяется расширенный код Хэмминга и записываются контрольные биты четности. Декодирование по алгоритму CLC может проходить в стандартном режиме (CLC-S) и расширенном (CLC-E). Расширенный режим, используя двухитерационный процесс корректирования ошибок, позволяет корректировать более сложные комбинации многократных ошибок [5]. Структура кодового слова CLC (16, 40) схематически представлена на рис. 1, где D – информационные биты, C – контрольные биты, P – биты проверки четности.

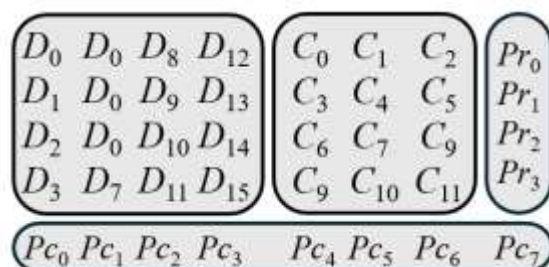


Рис. 1. Структура кодового слова алгоритма CLC (16, 40) [5]

Optimized Product Codes for Space Applications (OPCoSA) является дальнейшим развитием кода Product Codes for Space Applications (PCoSA), предложенный в источнике [7]. Алгоритм кодирования заключается в применении к 16-битной информационной матрице расширенного кода Хэмминга по строкам и столбцам, тем самым формируются контрольные биты и биты проверки четности для 4-х строк и столбцов [6]. В сравнении с кодом PCoSA, у которого кодовое слово состоит из 64 бит (16 бит информации и 48 проверочных бит), нововведением кода OPCoSA является сокращение кодового слова до 48 бит (рис. 2). Такие результаты были достигнуты при помощи удаления области повторного кодирования для уже полученных проверочных бит, тем самым уменьшив на 33 % общее количество проверочных бит. Несмотря на уменьшение количества проверочных бит, корректирующая способность показывает такую же эффективность для большинства комбинаций многократных ошибок [6].

D_0	D_1	D_2	D_3	Cl_0	Cl_1	Cl_2	Pl_0
D_4	D_5	D_6	D_7	Cl_3	Cl_4	Cl_5	Pl_1
D_8	D_9	D_{10}	D_{11}	Cl_6	Cl_7	Cl_8	Pl_2
D_{12}	D_{13}	D_{14}	D_{15}	Cl_9	Cl_{10}	Cl_{11}	Pl_3
$C2_0$	$C2_1$	$C2_2$	$C2_3$				
$C2_4$	$C2_5$	$C2_6$	$C2_7$				
$C2_8$	$C2_9$	$C2_{10}$	$C2_{11}$				
$P2_0$	$P2_1$	$P2_2$	$P2_3$				

Рис. 2. Структура кодового слова алгоритма OPCoSA (48, 16) [6]

Моделирование в программе Logisim

На первом этапе сравнительного анализа кодеров помехоустойчивого кодирования CLC и OPCoSA было проведено сравнение их фактической корректирующей способности многократных ошибок. Основываясь на предложенных в источниках [3 – 5] методах кодирования и декодирования, в программе Logisim были реализованы: кодер, модель канала передачи информации и декодер (рис. 3) кодеров CLC (16, 40) и OPCoSA (48, 16).

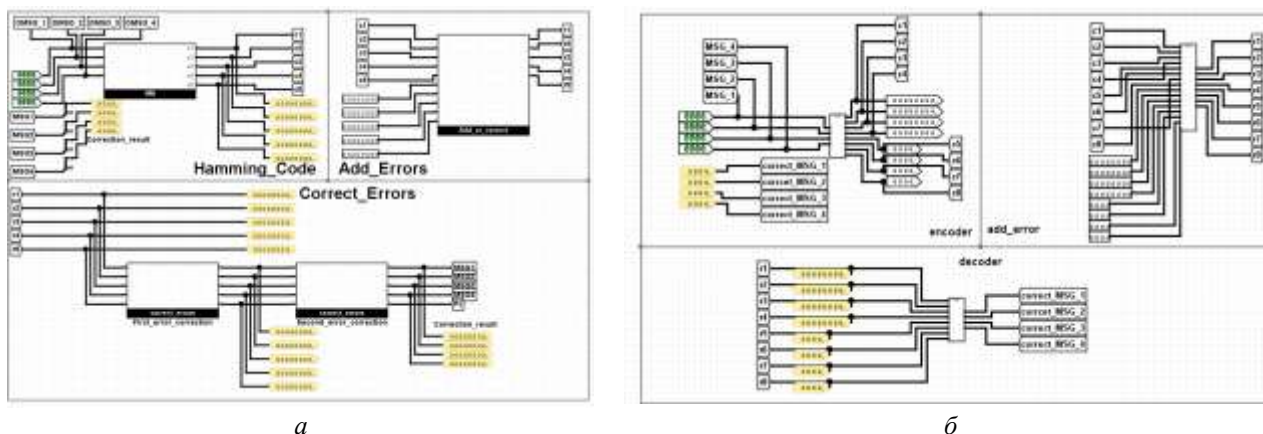


Рис. 3. Структурные схемы кодеров помехоустойчивого кодирования в программе Logisim: *а* – CLC; *б* – OPCoSA

Как показано на рис. 3, модули «Hamming_code» и «encoder» являются кодерами. Модуль «add_error» является моделью канала связи, которая позволяет вносить ошибки в кодовое слово. Модули «Correct_Errors» и «decoder» являются декодерами и способны корректировать ошибки в полученном из канала связи кодовом слове. Информация на входе кодера и информация на выходе декодера расположены рядом, для удобства сравнения.

Оценка корректирующей способности кодеров производилась путем внесения через модуль «add_errors» случайных комбинаций ошибок кратности от 1 до 8. Для каждого из проектов CLC и OPCoSA отдельно была применена симуляция по методу Монте-Карло. По итогам симуляции было определено количество успешно скорректированных ошибок для каждого алгоритма. Все тесты были проведены в одних и тех же условиях и при одинаковых комбинациях ошибок с целью сравнения фактической корректирующей способности кодеров CLC и OPCoSA.

По итогам моделирования построены диаграммы, представленные на рис. 4.

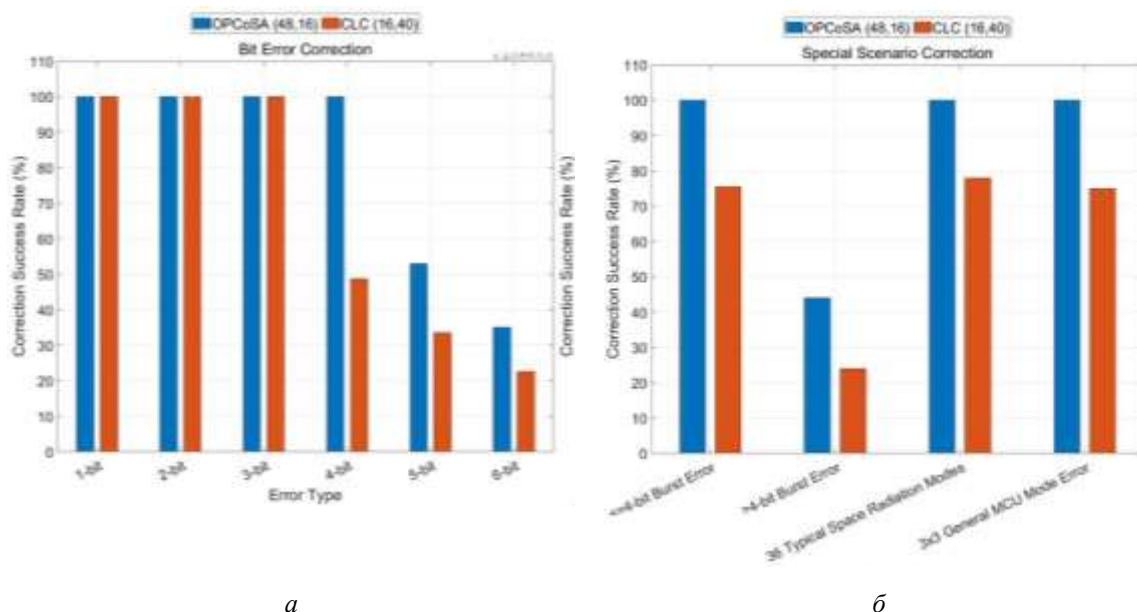


Рис. 4. Сравнение корректирующей способности CLC и OPCoSA: *а* – общий случай; *б* – специальные комбинации ошибок

Результаты экспериментов показывают, что оба алгоритма достигают 100 % корректирующей способности при возникновении от одной до трех ошибок. В случае четырех ошибок корректирующая способность алгоритма OPCoSA достигает 100 %, в то время как у CLC

она падает до 48, 69 %. При 5-кратных и 6-кратных ошибках, корректирующая способность для OPCoSA составляет 53 % и 35 % соответственно, что существенно больше, чем у алгоритма CLC, для которого данные показатели составляют 33,58 % и 22,58 %. При некоторых особых комбинациях ошибок преимущество OPCoSA еще более очевидно: корректирующая способность данного алгоритма составляет 100 % для пакетных ошибок кратности ≤ 4 и 36 типичных шаблонов ошибок в поле 3×3 , вызванных действием космической радиации. В свою очередь, при тех же условиях для CLC корректирующая способность варьируется от 75 % до 78 %. При пакетных ошибках кратности больше 4, OPCoSA также демонстрирует свое преимущество в корректирующей способности: для данного алгоритма возможна коррекция в 44 % случаев, в то время как у CLC данный показатель составляет 24 %. Обнаружено принципиальное различие между двумя алгоритмами, заключающееся в том, что коррекция ошибок CLC работает эффективно только в случае, если они располагаются в области 16-битного информационного поля, в то время как для OPCoSA представляется возможным корректировать ошибки вне зависимости от их положения в 48-битном кодовом слове. Это делает OPCoSA более предпочтительным для условий с частым возникновением многократных ошибок, так как для CLC изменение контрольных бит кодового слова приведет к неправильной коррекции ошибок.

Моделирование в САПР Xilinx Vivado

После симуляции кодеров в программе Logisim и оценки их корректирующих способностей, схема Logisim была переведена на язык описания аппаратуры Verilog, который впоследствии был импортирован в САПР Xilinx Vivado, представляющую собой среду для симуляции и программирования логических элементов ПЛИС фирмы Xilinx. В качестве платы для симуляции была использована Xilinx Vertex-7 серии XC7V485TFFG1157-1, так как она обладает высоким быстродействием и достаточным набором логических ресурсов для реализации алгоритмов CLC и OPCoSA на своей базе.

Дальнейшее моделирование проводилось с использованием инструментов САПР Xilinx Vivado. Для сравнительной оценки алгоритмов CLC и OPCoSA были получены данные об используемых платой ресурсах и потребляемой мощности. Результаты приведены на рис. 5.

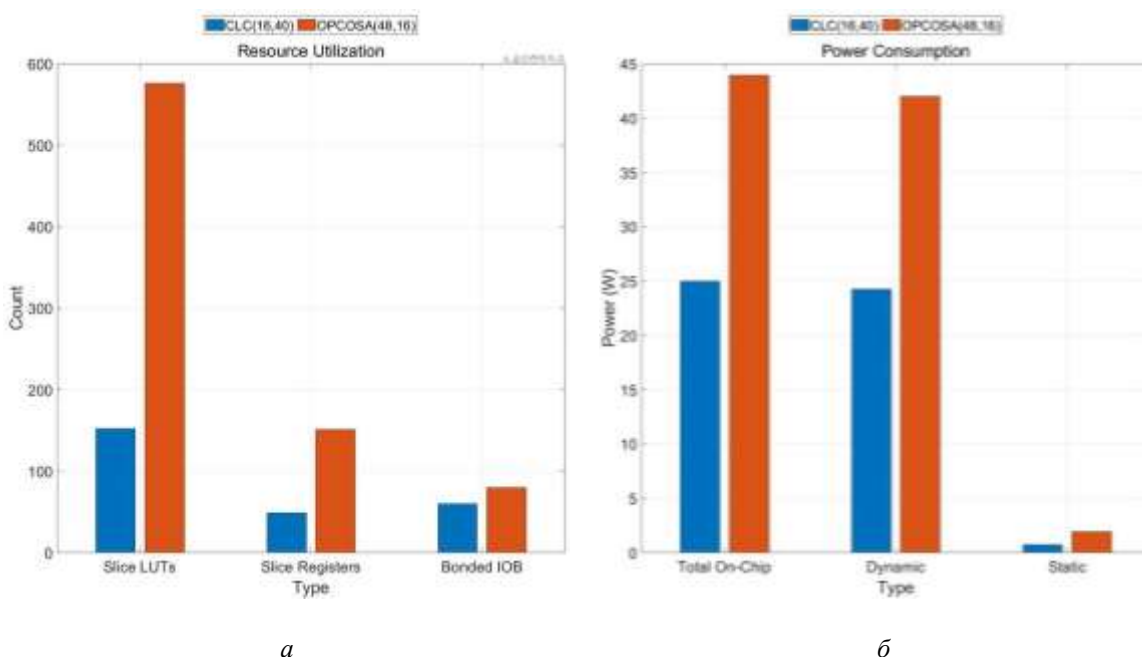


Рис. 5. Ресурсозатраты алгоритмов CLC и OPCoSA на ПЛИС:

а – использование логических ресурсов; *б* – энергопотребление

Результаты моделирования алгоритмов показали, что ресурсозатраты у OPCoSA больше чем у CLC приблизительно в 2 раза. Занятость LUT для алгоритма OPCoSA составила 1454 единиц, а регистров – 475 единиц. Для CLC эти показатели составляют 752 LUT и 251 регистров.

Энергопотребление демонстрирует аналогичное различие у обоих алгоритмов. У OPCoSA данный показатель составляет 7,6 Вт, в то время как энергопотребление алгоритма CLC на симулируемой плате составляет 5,2 Вт.

Заключение

Сравнительный анализ кодеков CLC (16, 40) и OPCoSA (48, 16) показал различие таких параметров как корректирующая способность и потребление логических ресурсов на ПЛИС. CLC подходит для устройств с ограниченными вычислительными ресурсами, которые будут применяться в областях с низкой вероятностью появления многократных ошибок. Несмотря на то, что OPCoSA представляет собой вариант, требующий больших логических ресурсов на ПЛИС, этот недостаток компенсируется достаточно высокой корректирующей способностью данного алгоритма. В условиях с большой вероятностью возникновения многократных ошибок, данный алгоритм позволяет создать гораздо более эффективное устройство передачи данных.

CLC AND OPCOSA: NOISE-RESISTANT CODING ALGORITHMS

T. A. TUMANOV, L. WEI, A. D. VASIUTICH, J. MA, X. REN, A.V. KURILOVICH

Abstract. A comparison was conducted between two-dimensional error-correction coding and decoding algorithms for information presented in a 4×4 matrix: Column-Line Codes (CLC) and Optimized Product Codes for Space Applications (OPCoSA). The algorithms were implemented using the Logisim digital logic simulator and the Vivado design suite for Xilinx Field-Programmable Gate Arrays (FPGAs). Quantitative assessments of the resources utilized by the codecs and the power consumption of the FPGAs were obtained.

Keywords: error-correcting codes, CLC, OPCoSA, modeling, multiple error combination.

Список литературы

1. Hazucha P., Svensson C. // IEEE Transactions on Nuclear Science. 2000. Vol. 47(6). P. 2586-2594.
2. Ibe E., Taniguchi H., Yahagi Y., Shimbo K., Toba T // IEEE Trans. Electron Devices. 2010. Vol. 57(7). P. 1527-1538.
3. Argyrides C., Zarandi H., Pradhan D. // Proc. IEEE Int. Symp. Defect Fault-Tolerance VLSI Syst. (DFT). 2007. P. 340-348.
4. Castro H., Silveira J., Coelho A., Silva F., Magalhaes P., Lima O. // Proc. IEEE Int. New Circuits Syst. Conf. (NEWCAS). 2016. P. 1-4.
5. Silva F., Silveira J., Marcon C., Vargas F., Lima O. Jr. // Journal of Electronic Testing. 2018. Vol. 34(4). P. 417-433.
6. Freitas D., Silveira J., Marcon C., Naviner L., Mota J. // Integration. 2022. Vol. 84. P. 131-141.
7. Freitas D., Mota D., Goerl R., Marcon C., Vargas F., Silveira J., Mota J. // Integration. 2020. Vol. 74. P. 71-80.

ПРОЕКТ ЕДИНОЙ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ (ЕАП) ДЛЯ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ИНФОКОММУНИКАЦИОННОЙ КОМПАНИИ: ПЕРЕХОД К УПРАВЛЕНИЮ, ОСНОВАННОМУ НА ДАННЫХ, В РУП «БЕЛТЕЛЕКОМ»

Д.М. ГОРБУКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 17 апреля 2026

Аннотация. В статье обоснована необходимость замены ручной табличной обработки данных на сквозную аналитическую платформу для РУП «Белтелеком». Описаны текущие проблемы: отсутствие единой витрины данных, высокая трудоёмкость отчётности, невозможность глубокого анализа, риски несогласованности данных. Предложена архитектура решения на базе Open Source компонентов (Kafka, ClickHouse, Airflow, Superset, Kubernetes, OpenMetadata, Keycloak). Показаны ожидаемые выгоды: сокращение времени подготовки отчётов до 80%, снижение ошибок на 90%, ускорение принятия решений, возможность прогнозной аналитики оттока и LTV.

Ключевые слова: цифровая трансформация, управление данными, телеком-аналитика, data platform, Open Source, ClickHouse, Apache Airflow, бизнес-интеллект, качество данных.

Введение

Ключевая проблема: система управления и анализа данных, построенная на ручной обработке с использованием табличных методов, не соответствует масштабам и динамике бизнеса компании. Это создает критические риски для операционного управления, финансовых результатов и конкурентной позиции РУП «Белтелеком».

Невозможность качественного drill-down-анализа: детализация (например, динамика ARPU с помесечной детализацией по каждой услуге и с детализацией по филиалам и ЗУЭС, УЭС) в рамках ручных методов приводит к созданию чрезвычайно сложных и хрупких моделей. Любая попытка «копнуть глубже» требует создания нового ручного отчета.

Слепые зоны в клиентском опыте. При абонентской базе в 2,5 млн пользователей интернета byfly и 2,53 млн зрителей TV ZALA, ручной анализ оттока и причин недовольства неэффективен. Это ведет к потере лояльных клиентов. Что особенно актуально в жесткой конкурентной среде, в которой работает компания.

Ручной труд и высокая трудоемкость. Большинство отчетов предположительно для управления коммерческой деятельностью и руководства формируются вручную с использованием формул (ВПР, расчетные столбцы и др. – распространённая методика подготовки таких отчетов), что требует значительных временных затрат и подвержено ошибкам.

Разрозненность данных. Данные извлекаются из разрозненных источников, что усложняет консолидацию и верификацию информации.

Низкая оперативность. Ежедневные отчеты готовятся с задержками из-за необходимости ручного сбора данных.

Сложность адаптации к изменениям. При изменении структуры филиалов (например, объединение финрегионов в Гомельском филиале) или критериев отчетности (например, новые тарифные планы) требуются ручные доработки формул и макросов. Постоянные доработки отчетов и ручные корректировки требуют вовлечения сотрудников.

Риски несогласованности данных. Отчеты из разных источников могут содержать противоречивые данные из-за различий в методологиях расчета.

Риски упустить выгоду от кросс-продаж. «Белтелеком» предлагает десятки услуг и пакетов («ЯСНА», «Бизнес-кейс», byfly, ZALA, «Видеоконтроль»), но без единой витрины данных сложно выявлять абонентов, готовых к апгрейду тарифа или подключению дополнительных услуг (например, к пакету «Видеоконтроль»).

Вопросы оперативного управления скидками при ценообразовании: компания активно использует региональные и стимулирующие скидки (до 60% на 18 месяцев). Без оперативного анализа их влияния на LTV (Lifetime Value) и отток клиентов невозможно оценить их истинную рентабельность и оптимизировать структуру.

В настоящей статье предложена архитектура сквозной аналитической платформы на Open Source стеке, позволяющая перейти от реактивного управления к проактивному, основанному на данных.

Текущее состояние, потенциал

«Белтелеком» обладает огромным массивом данных, но испытывает дефицит управляемой аналитики.

Масштаб данных: миллионы абонентов. Белтелеком – крупный игрок с 2,5 млн. интернет-абонентов и 2,53 млн. подписчиков ТВ. При этом у компании сотни тарифных комбинаций и множество сервисов (Evika, «Мой город» и др.), сложная тарифная структура с пакетами «ЯСНА», региональными различиями и программами лояльности.

Технические системы по предоставлению услуг (оборудование сети передачи данных, коммуникационные системы, сервисные платформы, билинговая система и др.) генерируют огромный объем операционной информации.

Ограниченная аналитика. Отсутствуют данные об внутренних системах аналитики. При этом в каждом сегменте услуг Белтелекома или региона страны присутствуют активные конкуренты. Принятие решений часто основано на агрегированных или устаревших отчетах, что не позволяет проводить глубокий анализ в разрезе отдельных сегментов.

Подтвержденная потребность. Успешный запуск платформы «Мой город» и активное развитие экосистемы услуг (от зарядных станций до видеонаблюдения) свидетельствует о том, что компания готова к внедрению сложных IT-решений для улучшения клиентского сервиса и операционной эффективности.

Проблемы:

- отсутствие единой платформы для управления данными;
- отсутствие автоматизированной верификации данных.

Закключение: текущая модель ручной обработки данных не является устойчивой. Она блокирует рост, маскирует реальные проблемы и препятствует переходу к управлению, основанному на данных, что ставит под угрозу конкурентные преимущества компании.

Предложения по решению и ключевые технологии

Для автоматизации отчетности предлагается внедрить сквозную платформу анализа данных, охватывающую весь жизненный цикл – от интеграции источников до визуализации.

Интеграция и обработка данных. Автоматизировать сбор данных из всех источников.

ETL/ELT-процессы для автоматизированного сбора данных из всех источников в реальном времени и их очисткой. Например, оборудование сети передачи данных, коммуникационные системы, сервисные платформы, билинговая система, системы учета абонентов и услуг, системы расчетов, системы технического учета и паспортизации, бухгалтерской системы и др. Реализовать ETL/ELT-процессы предлагается на базе Apache Airflow [7, 8, 9] для оркестрации загрузки данных из всех источников.

Apache Kafka и Apache Spark Streaming для потоковой обработки [1, 2, 3]:

- новые подключения услуг в реальном времени;
- оперативные данные по продажам;
- мгновенное обновление дашбордов.

Data Lake для консолидации сырых данных из всех систем. Единое хранилище для сырых данных с возможностью очистки и стандартизации. Пример: данные о населении регионов, блокировках абонентов, доходах.

Data Lake (например, на базе MinIO или HDFS) для консолидации сырых данных из всех систем.

Great Expectations [16, 17] и dbt [21] для автоматических проверок качества данных (полнота, уникальность, отсутствие аномалий).

Выгода:

- исключение ручного переноса данных между системами;
- сокращение времени на подготовку отчетов на 70–80%;
- автоматическое обнаружение проблем с данными.

Хранение и обработка. Гибридное хранилище данных: создать единое надежное хранилище для исторических и оперативных данных.

Использовать аналитическое Data Warehouse [4, 5, 6] как основной слой для агрегатов и витрин, а ClickHouse, как колоночная СУБД, оптимизированная для телеком-аналитики, имеющая высокую скорость агрегаций и низкую стоимость хранения.

Реляционное хранилище для метаданных, справочников и управленческой отчетности построить на PostgreSQL и Patroni:

- стандартизированные бизнес-метрики;
- автоматический пересчет градаций районов;
- единая методология расчетов (например, ARPU, приросты и т.д.).

Kubernetes [13, 14] для оркестрации контейнеров (Docker) с компонентами платформы, что обеспечит масштабируемость и отказоустойчивость.

Выгода:

- накопление единой версии данных для всех филиалов;
- мгновенный доступ к истории изменений (например, миграция регионов между группами);
- избежать эффекта «хрустального сервера»;
- высокая производительность даже при многолетней истории и миллионах абонентов.

Управление метаданными и качеством данных.

Data Catalog: OpenMetadata [15] или DataHub – для поиска данных, управления метаданными, lineage (происхождения данных).

Data Quality: Great Expectations [16, 17] (автоматические проверки), dbt tests [21], мониторинг аномалий.

Data Observability: метрики здоровья пайплайнов, оповещения о сбоях.

Выгода:

- прозрачность происхождения данных;
- снижение времени на поиск нужных данных;
- доверие к данным со стороны бизнеса.

Визуализация и аналитика. Обеспечить руководство интуитивными отчетами и дашбордами для ежедневного контроля.

Внедрение платформы Apache Superset [10, 11, 12] для создания интерактивных дашбордов и отчетов:

– Дашборды для тактического управления (управления коммерческой деятельностью): использование интерактивных фильтров и возможностей drill-down для самостоятельного анализа динамики продаж, выполнения планов и прогнозных моделей в реальном времени.

– Стратегические отчеты (руководство компании): создание единой панели KPI с использованием богатой библиотеки визуализаций Superset для отображения трендов и паттернов.

Настроить автоматические уведомления о рисках (снижение ARPU, рост блокировок).

Настроить API для интеграции с другими системами компании.

Выгода:

- сокращение времени формирования отчета с часов до минут;
- возможность drill-down до уровня филиала; точность данных за счет исключения ручных ошибок.

Безопасность.

Keycloak [18, 19] или Apache Ranger для централизованной аутентификации и авторизации.

Шифрование данных в покое (Data Lake, DWH) и при передаче (TLS).

Аудит доступа к данным и отчётам.

Маскирование чувствительных данных (персональные данные абонентов).

Интеграция с корпоративным LDAP/Active Directory.

Выгода:

- соответствие законодательству РБ о персональных данных;
- защита от утечек;
- прозрачность доступа.

Оркестрация и DevOps.

Apache Airflow [7, 8, 9] – оркестрация ETL/ELT-пайплайнов.

GitLab CI/CD или Jenkins для автоматического развертывания.

Terraform или Ansible (Infrastructure as Code).

Prometheus и Grafana для мониторинга платформы.

Выгода:

- воспроизводимость инфраструктуры;
- быстрое развёртывание новых окружений;
- высокая надёжность.

Прогнозная аналитика и Reverse ETL.

ML-платформа (например, на базе JupyterHub и MLflow) для:

- прогноза оттока (churn prediction),
- прогноза ARPU и LTV,
- рекомендательных систем для кросс-продаж.

Reverse ETL (инструменты типа Hightouch, Census или самописные конвейеры) для выгрузки сегментов и инсайтов обратно в CRM и биллинг.

Выгода:

- проактивное удержание клиентов;
- автоматические персонализированные предложения;
- рост LTV – пожизненной ценности клиента, упрощённая формула для телекома:

$LTV = (\text{Средний платёж в месяц} \times \text{Средняя продолжительность жизни клиента в месяцах}) - \text{Расходы на привлечение и обслуживание.}$

Ожидаемые выгоды от реализации проекта

Сокращение времени подготовки отчетов:

- До 80% за счет автоматизации (сейчас до 4-6 часов в день на ручные операции).
- Формирование отчетов с 2 дней до 15 минут. Доступ к данным в реальном времени.
- Сокращение цикла планирования на 40%.
- Мгновенное добавление новых услуг (1-2 дня вместо 2-3 недель).

Повышение точности данных:

- Единая версия данных для всех отчетов.
- Исключение ручных ошибок через единые алгоритмы расчета (например: ARPU по всем видам услуг; проникновение услуг с учетом специфики регионов таких как количество и плотность населения, доходы населения, наличия конкурентов в разных сегментах услуг и др.).
- Снижение ошибок данных на 90%.

Ускорение принятия решений:

- Формирование отчета в реальном времени (вместо ежедневных усилий).
- Ускорение принятия решений на 50%.
- Доступ к актуальным данным в реальном времени для анализа эффективности филиалов.

Операционная эффективность:

- Масштабируемость для бизнес-роста.
- Возможность быстрого добавления новых метрик (например, метрики для услуги «Окко» или «Литрес»).
- Возможность прогнозной аналитики.

- Повышение прозрачности процессов на 90%.
 - Ускорение принятия стратегических решений.
- Снижение операционных рисков:
- Централизованный контроль данных и автоматическое логирование изменений.

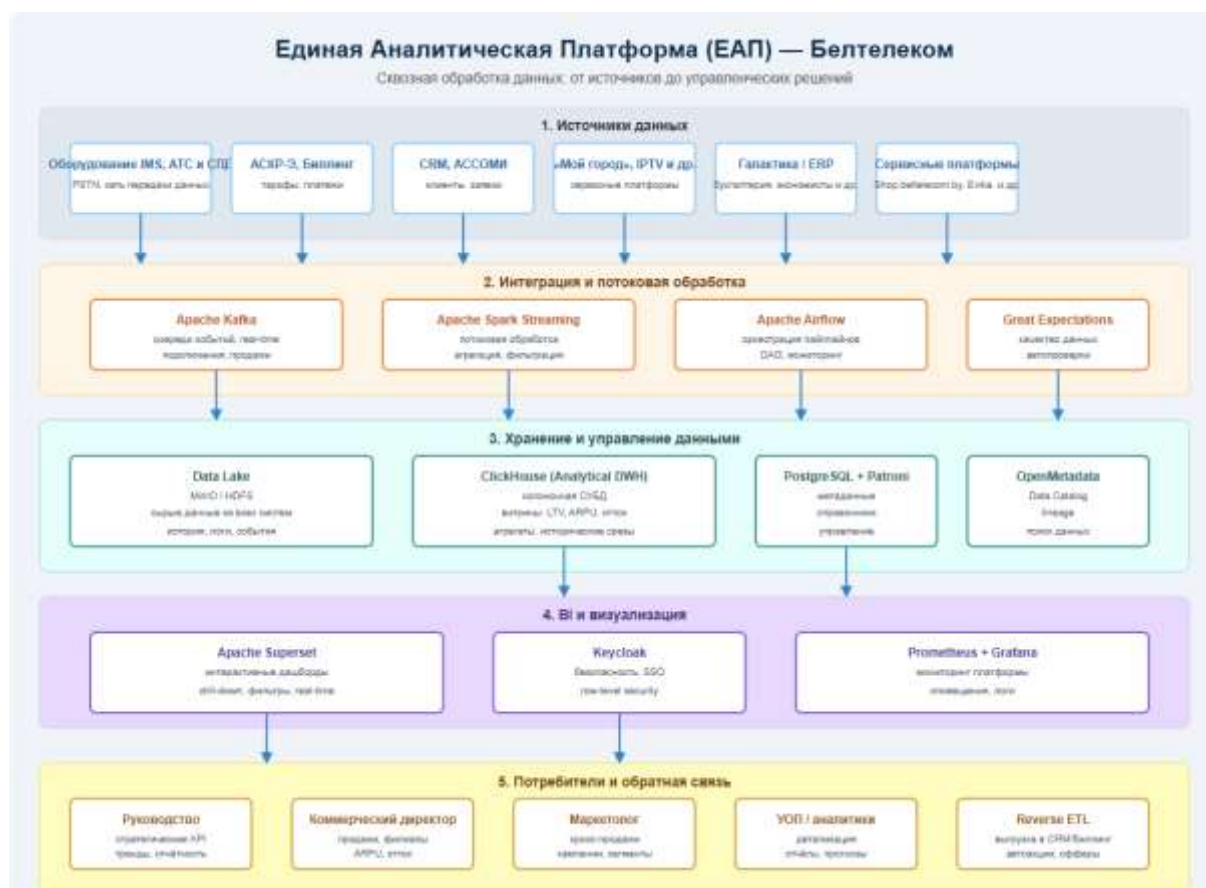


Рис. 1. Единая Аналитическая Платформа – Open Source стек для цифровой трансформации РУП «Белтелеком»

Заключение

Внедрение предложенной платформы позволит РУП «Белтелеком» совершить переход от фрагментированной ручной отчетности к единой управляемой экосистеме данных. Ключевые измеримые результаты:

- сокращение времени подготовки отчетов с 2 дней до 15 минут (*экономия до 80% трудозатрат*);
- снижение ошибок данных на 90% *за счёт автоматической верификации*;
- рост LTV и снижение оттока благодаря прогнозной аналитике и персонализированным предложениям.

Платформа построена на современном Open Source стеке, что обеспечивает низкую совокупную стоимость владения, масштабируемость и независимость от вендоров. Пилотный проект минимизирует риски, а поэтапное масштабирование позволяет контролировать затраты на каждом шаге.

Реализация данного решения создаст основу для цифровой трансформации управления данными и перехода к data-driven культуре, что напрямую повысит конкурентоспособность компании на рынке телекоммуникационных услуг.

UNIFIED ANALYTICS PLATFORM (UAP) PROJECT FOR THE DIGITAL TRANSFORMATION OF AN INFOCOMMUNICATION COMPANY: TRANSITION TO DATA-DRIVEN MANAGEMENT AT RUP «BELTELECOM»

D.M. HARBUKOU

Abstract. The document substantiates the need to replace manual spreadsheet-based data processing with an end-to-end analytics platform for RUP «Beltelecom». Current problems are described: lack of a single data mart, high reporting labor intensity, inability to perform deep analysis, and risks of data inconsistency. A solution architecture based on Open Source components (Kafka, ClickHouse, Airflow, Superset, Kubernetes, OpenMetadata, Keycloak) is proposed. Expected benefits include up to 80% reduction in reporting time, 90% reduction in data errors, faster decision-making, and predictive analytics for churn and LTV.

Keywords: digital transformation, data management, telecom analytics, data platform, Open Source, ClickHouse, Apache Airflow, business intelligence, data quality.

Список литературы

1. Apache Kafka [Электронный ресурс] // Официальный сайт Apache Kafka. – Режим доступа: <https://kafka.apache.org/>. – Дата доступа: 17.04.2026.
2. Kafka in Telecom: реальные кейсы использования [Электронный ресурс] // Confluent. – Режим доступа: <https://www.confluent.io/resources/customer-stories/telecom/>. – Дата доступа: 17.04.2026.
3. Архитектура Apache Kafka (диаграммы) [Электронный ресурс] // Apache Kafka Documentation. – Режим доступа: <https://kafka.apache.org/24/images/kafka-apis.png>. – Дата доступа: 17.04.2026.
4. ClickHouse [Электронный ресурс] // Официальный сайт ClickHouse. – Режим доступа: <https://clickhouse.com/>. – Дата доступа: 17.04.2026.
5. ClickHouse для телеком-решений: примеры использования [Электронный ресурс] // ClickHouse. – Режим доступа: <https://clickhouse.com/use-cases/telecom>. – Дата доступа: 17.04.2026.
6. Архитектура ClickHouse: колоночное хранение, партиционирование, сжатие [Электронный ресурс] // ClickHouse Documentation. – Режим доступа: <https://clickhouse.com/docs/en/development/architecture>. – Дата доступа: 17.04.2026.
7. Apache Airflow [Электронный ресурс] // Официальный сайт Apache Airflow. – Режим доступа: <https://airflow.apache.org/>. – Дата доступа: 17.04.2026.
8. Архитектура Apache Airflow (схема компонентов) [Электронный ресурс] // Apache Airflow Documentation. – Режим доступа: https://airflow.apache.org/docs/apache-airflow/stable/_images/airflow_architecture.png. – Дата доступа: 17.04.2026.
9. Astronomer. Data Engineering with Airflow: практические руководства [Электронный ресурс] // Astronomer. – Режим доступа: <https://www.astronomer.io/guides/>. – Дата доступа: 17.04.2026.
10. Apache Superset [Электронный ресурс] // Официальный сайт Apache Superset. – Режим доступа: <https://superset.apache.org/>. – Дата доступа: 17.04.2026.
11. Галерея дашбордов Apache Superset [Электронный ресурс] // Apache Superset Gallery. – Режим доступа: <https://superset.apache.org/gallery/>. – Дата доступа: 17.04.2026.
12. Архитектура Apache Superset [Электронный ресурс] // Apache Superset Documentation. – Режим доступа: <https://superset.apache.org/docs/contributing/architecture>. – Дата доступа: 17.04.2026.
13. Kubernetes. Компоненты кластера [Электронный ресурс] // Kubernetes Documentation. – Режим доступа: <https://kubernetes.io/docs/concepts/overview/components/>. – Дата доступа: 17.04.2026.
14. Kubernetes в телеком-индустрии: примеры внедрения [Электронный ресурс] // Cloud Native Computing Foundation (CNCF). – Режим доступа: <https://www.cncf.io/case-studies/telecom/>. – Дата доступа: 17.04.2026.
15. OpenMetadata. Архитектура платформы [Электронный ресурс] // OpenMetadata. – Режим доступа: <https://open-metadata.org/architecture>. – Дата доступа: 17.04.2026.
16. Great Expectations [Электронный ресурс] // Официальный сайт Great Expectations. – Режим доступа: <https://greatexpectations.io/>. – Дата доступа: 17.04.2026.
17. Great Expectations Gallery: примеры отчётов о качестве данных [Электронный ресурс] // Great Expectations. – Режим доступа: <https://greatexpectations.io/gallery>. – Дата доступа: 17.04.2026.
18. Keycloak. Архитектура и основные концепции [Электронный ресурс] // Keycloak Documentation. – Режим доступа: https://www.keycloak.org/docs/latest/server_admin/#core-concepts-and-terms. – Дата доступа: 17.04.2026.

19. Keycloak. Интеграция с BI-инструментами [Электронный ресурс] // Keycloak Guides. – Режим доступа: <https://www.keycloak.org/guides>. – Дата доступа: 17.04.2026.
20. AWS for Telecom. Референтная архитектура платформы данных [Электронный ресурс] // Amazon Web Services. – Режим доступа: <https://aws.amazon.com/telecom/data-platform/>. – Дата доступа: 17.04.2026.
21. dbt Labs. Современный стек данных в картинках [Электронный ресурс] // dbt Blog. – Режим доступа: <https://www.getdbt.com/blog/modern-data-stack-in-pictures/>. – Дата доступа: 17.04.2026.
22. Databricks. Архитектура Data Lakehouse для телекома [Электронный ресурс] // Databricks. – Режим доступа: <https://www.databricks.com/solutions/industries/telecommunications>. – Дата доступа: 17.04.2026.

СРАВНЕНИЕ АЛГОРИТМОВ ПОДБОРА ВЕСОВОГО КОЭФФИЦИЕНТА ОБЪЕДИНЕНИЯ ИЗОБРАЖЕНИЙ АТОМНО-СИЛОВОЙ МИКРОСКОПИИ

М.Ю. ЛОВЕЦКИЙ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 24 апреля 2026

Аннотация. В работе решается задача ускорения подбора весового коэффициента при взвешенном сложении многоканальных изображений атомно-силовой микроскопии на основе локальной корреляционной метрики. Предложен алгоритм, исключающий этап сегментации за счёт прямого поиска оптимума от центра диапазона весов, что на выборке из 385 пар изображений обеспечило ускорение в 6 раз относительно глобального перебора при статистически незначимом отклонении коэффициента.

Ключевые слова: атомно-силовая микроскопия, взвешенное сложение каналов, локальная корреляционная метрика, ускоренная обработка изображений, бинаризация Оцу.

Введение

Многоканальная атомно-силовая микроскопия (АСМ) обеспечивает одновременную регистрацию нескольких физических характеристик поверхности, таких как топография, жёсткость и вязкость, формируя набор синхронизированных одноканальных изображений. Для повышения информативности визуального анализа применяется процедура взвешенного сложения каналов, позволяющая синтезировать единое комбинированное изображение с сохранением наиболее значимых локальных деталей каждого исходного канала [1,2]. Ключевым параметром данного преобразования выступает весовой коэффициент, определяющий относительный вклад каждого канала в итоговую интенсивность пикселей. Некорректный выбор веса приводит к потере контраста, сглаживанию микрорельефа или усилению шумовых артефактов, что снижает диагностическую ценность полученных данных. Оценка качества объединения осуществляется с помощью локальной корреляционной метрики [3], вычисляемой в скользящем квадратном окне. Данная метрика учитывает пространственные взаимосвязи между каналами и демонстрирует высокую чувствительность к сохранению локальных неоднородностей, в отличие от других известных метрик качества [4-7], однако её значения существенно зависят от размера анализирующего окна. Согласование размера окна с преобладающими масштабами структур на поверхности является необходимым условием корректной оценки, что в классических реализациях достигается путём полного перебора комбинаций весового коэффициента и размеров окна. Такой подход гарантирует нахождение глобального оптимума, но характеризуется высокой вычислительной сложностью, делающей его неприменимым для обработки больших массивов данных или систем реального времени. Необходимость оптимизации вычислительной нагрузки привела к разработке сегментных подходов, в которых размер окна определяется на основе предварительного анализа пространственной структуры изображений. В частности, применение бинаризации по методу Оцу с последующим построением гистограммы эффективных размеров выделенных областей позволяет зафиксировать размер окна по модальному значению распределения, существенно сокращая пространство поиска. Несмотря на доказанную эффективность, сегментная обработка вводит дополнительные этапы предобработки, увеличивает длительность вычислений и требует контроля параметров бинаризации для избежания артефактов на границах разнородных областей. Возникает научная и практическая необходимость в разработке алгоритма, способного

обеспечить сопоставимую точность подбора весового коэффициента без использования процедур сегментации и статистического анализа размеров областей. Целью настоящей работы является повышение точности и скорости подбора весового коэффициента сложения информационных каналов изображений АСМ.

Материалы и методы

Исследование проводилось на наборе данных, включающем 385 пар синхронизированных АСМ-изображений, полученных в различных режимах сканирования (пример на рисунке 1).

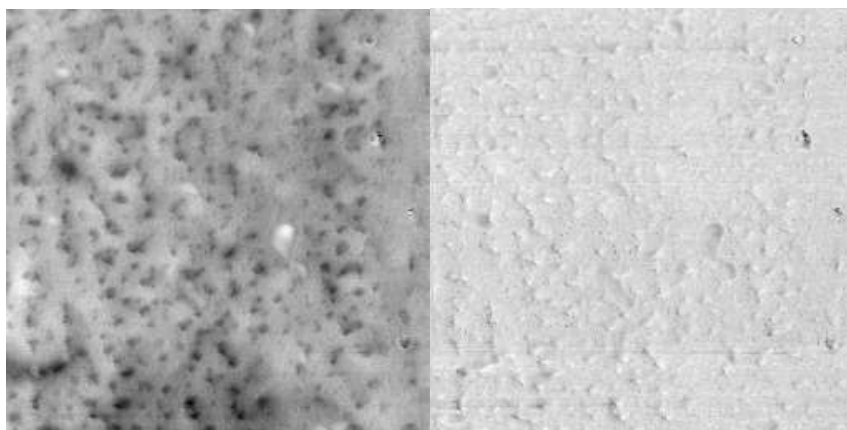


Рисунок 1. Пример двух каналов АСМ-изображений

Синтез комбинированного изображения осуществлялся посредством взвешенного сложения пиксельных значений двух исходных каналов в соответствии с выражением

$$m_c(y, x) = [k \cdot m_1(y, x) + (1 - k) \cdot m_2(y, x)], \quad (1)$$

где $m_1(y, x)$ и $m_2(y, x)$ представляют собой интенсивности пикселей первого и второго каналов соответственно, k является весовым коэффициентом, принимающим значения в интервале от нуля до единицы, $m_c(y, x)$ обозначает результирующую интенсивность комбинированного изображения.

Физический смысл коэффициента k заключается в балансировке вклада каналов таким образом, чтобы минимизировать искажения значимых локальных структур при слиянии. Оценка качества объединения выполнялась с использованием локальной корреляционной метрики $D_L(p)$, вычисляемой по формуле

$$D_L(p) = \left[\frac{r_L(M_c, M_1, p) + r_L(M_c, M_2, p)}{|r_L(M_c, M_1, p) - r_L(M_c, M_2, p)| \cdot r_L(M_1, M_2, 0.5)} \right] \quad (2)$$

где r_L обозначает коэффициент локальной корреляции между соответствующими изображениями в квадратном окне диаметром в p пикселей.

Метрика учитывает взаимную корреляцию комбинированного изображения с каждым исходным каналом, а также корреляцию между самими каналами, что обеспечивает комплексную оценку сохранности информации. Чем выше значение $D_L(p)$, тем лучше качество объединения. В работе сравнивались три алгоритмических подхода к определению оптимального значения k .

Эталонный метод реализует полный перебор коэффициента k в диапазоне от нуля до единицы и размера окна p в широком интервале возможных значений. Для каждой комбинации параметров вычисляется метрика $D_L(p)$, после чего выбирается пара, обеспечивающая

максимальное значение. Данный подход гарантирует нахождение глобального оптимума, однако его вычислительная сложность пропорциональна произведению количества шагов по k и p , что приводит к значительным временным затратам.

Сегментный метод снижает размерность задачи за счёт предварительного определения оптимального размера окна. На первом этапе выполняется бинаризация каждого компонентного изображения по алгоритму Оцу, разделяющая пиксели на объектную и фоновую области. Далее строится гистограмма эффективных размеров выделенных сегментов, и её модальное значение принимается за стартовый размер окна p , на основе которого осуществляется поиск методом золотого сечения размера окна p и весового коэффициента k с заданным шагом до достижения максимума $D_L(p)$.

Предлагается метод, полностью исключаящий этапы бинаризации, сегментации и построения гистограмм. Анализ зависимостей $D_L(p)$ для различных пар изображений показал, что функция метрики обладает гладким характером и содержит единственный ярко выраженный максимум в исследуемом диапазоне. Это свойство позволяет производить поиск оптимального веса непосредственно от центров диапазона $k = 0,5$; $p = 62$. Перебор осуществляется пошагово с фиксированным шагом $\Delta k = 0,033$; $\Delta p = 1$ в обе стороны от начальной точки до тех пор, пока значение метрики не начнёт убывать, что свидетельствует о прохождении экстремума. Шаги были выбраны исходя из баланса между точностью определения оптимума и доступными временными ресурсами. Статистическая значимость различий между методами оценивалась с применением парного t -критерия Стьюдента при уровне значимости $\alpha = 0,05$. Все вычисления выполнялись в идентичных программно-аппаратных условиях для обеспечения воспроизводимости результатов.

Результаты

После применения предложенного метода совмещения информационных каналов изображения, результирующее изображение несет в себе больше данных для дальнейшего анализа (пример комбинированного изображения приведен на рисунке 2).

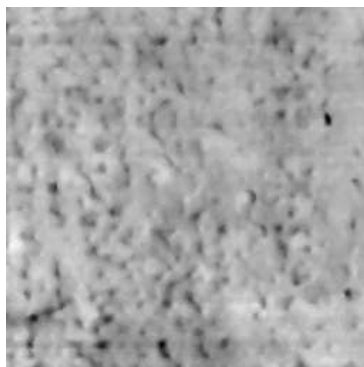


Рис 2. Пример объединения двух каналов изображения

Сравнительный анализ вычислительной эффективности трёх рассмотренных подходов выявил существенные различия во временных затратах на обработку одной пары изображений. Среднее время выполнения эталонного глобального перебора составило $2,80 \pm 0,05$ с, что отражает высокую стоимость полного сканирования пространства параметров. Применение сегментного метода на основе бинаризации Оцу и анализа моды гистограммы позволило сократить среднее время обработки до $0,47 \pm 0,01$ с. Предложенный алгоритм, исключая этап сегментации, продемонстрировал среднее время выполнения $0,46 \pm 0,01$ с. Относительно эталонного подхода ускорение обоих методов достигло 6 раз, что наглядно иллюстрируется на рисунке 3.

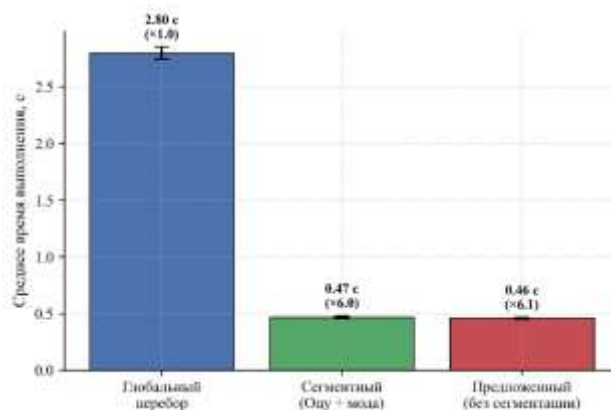


Рис. 3. Сравнение среднего времени выполнения методов подбора весового коэффициента

Разница во времени выполнения между сегментным и предложенным методами составила $1,63\% \pm 3,21\%$, что по результатам парного t-критерия Стьюдента является статистически незначимым.

Точность определения весового коэффициента оценивалась по величине отклонения найденного значения k от эталонного оптимума, полученного глобальным перебором. Для предложенного метода среднее отклонение составило $0,037 \pm 0,076$, тогда как для сегментного подхода данный показатель равен $0,040 \pm 0,075$. Распределение отклонений для обоих ускоренных методов демонстрирует близкие статистические характеристики и представлено на рисунке 4.

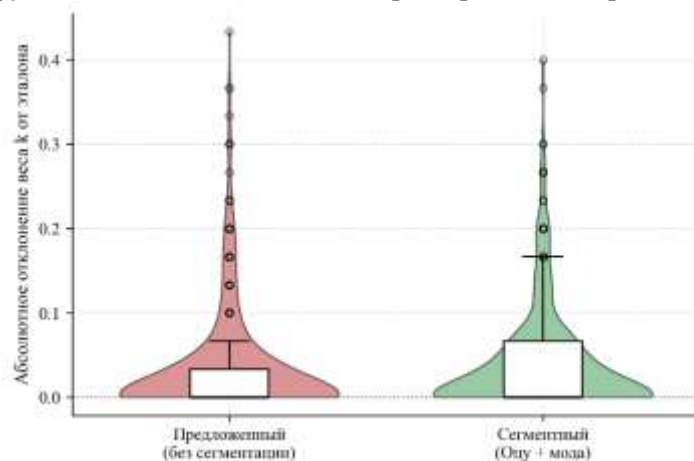


Рис. 4. Распределение отклонений найденного весового коэффициента k от эталонного значения

Статистический анализ подтвердил отсутствие значимых различий в точности подбора веса между ускоренными подходами ($p > 0,05$). Дополнительный анализ форм зависимостей $D_L(p)$ для всех 385 пар изображений показал, что кривые метрики характеризуются высокой гладкостью и наличием единственного глобального максимума при любых допустимых значениях p . Типичный пример зависимостей представлены на рисунке 5.

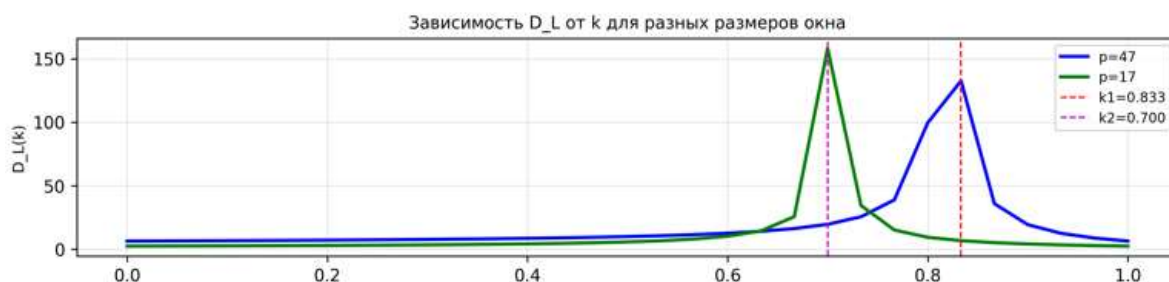


Рис. 5. Типичные зависимости локальной корреляционной метрики $D_L(p)$ от весового коэффициента k

Отсутствие локальных экстремумов и монотонное поведение функции в окрестностях оптимума объясняют устойчивость результатов предложенного метода к выбору начальной точки поиска и подтверждают корректность стратегии центрированного перебора.

Обсуждение

Полученные результаты свидетельствуют о том, что исключение этапа сегментации и анализа гистограмм размеров областей не приводит к ухудшению точности подбора весового коэффициента при взвешенном сложении АСМ-изображений. Теоретическое обоснование данного факта заключается в специфических свойствах локальной корреляционной метрики $D_L(p)$. Как показано в разделе результатов, зависимость метрики от весового коэффициента имеет гладкий, унимодальный характер на всём исследуемом интервале. Это означает, что функция не содержит локальных экстремумов, способных нарушить работу алгоритмов локального поиска, а её градиент устойчиво направлен в сторону глобального оптимума. В таких условиях стратегия инициирования перебора от центра диапазонов $k = 0,5, p = 62$ оказывается математически обоснованной, поскольку начальная точка с высокой вероятностью попадает в область монотонного роста метрики, обеспечивая быструю сходимость к максимуму. Стойкость метрики к вариациям начальных условий дополнительно усиливается за счёт усредняющего эффекта локального корреляционного окна, которое сглаживает высокочастотные шумы и стабилизирует оценку качества объединения. Практическая ценность предложенного подхода определяется существенным упрощением вычислительного процесса. Отказ от бинаризации по Оцу, построения гистограмм и вычисления модальных значений устраняет накладные расходы на предварительную обработку, которые в сегментном методе могут достигать 15–20% от общего времени выполнения. На больших выборках данных или в системах, требующих обработки изображений в режиме реального времени, такое упрощение напрямую транслируется в повышение пропускной способности и снижение задержек. Кроме того, устранение этапа сегментации снимает проблему чувствительности алгоритма к параметрам бинаризации и артефактам на границах неоднородных областей, что повышает стабильность работы на изображениях со сложной текстурой или низким контрастом. Тем не менее, предложенный метод имеет определённые ограничения. В случаях, когда распределение размеров структур на поверхности носит резко многомодальный характер, фиксированный шаг перебора $\Delta k = 0,033$ может оказаться недостаточно адаптивным, а расположение начальной точки поиска потребует дополнительных итераций для достижения оптимума. Для таких сценариев перспективным направлением является внедрение адаптивного шага, динамически изменяющегося в зависимости от локального градиента метрики, либо гибридная стратегия, комбинирующая быстрый центрированный поиск с выборочной проверкой границ диапазона. Перспективы дальнейшего развития алгоритма связаны с его интеграцией в специализированное программное обеспечение для автоматизированной обработки многоканальных АСМ-данных. Расширение метода на случай слияния более двух каналов потребует модификации метрики $D_L(p)$ и перехода к многомерной оптимизации весовых коэффициентов, однако свойство гладкости целевой функции, вероятно, сохранится, что позволит применить аналогичные стратегии ускоренного поиска. Дополнительно представляет интерес тестирование алгоритма на данных других видов сканирующей зондовой микроскопии, где задачи взвешенного объединения каналов также являются актуальными.

Заключение

В работе проведено сравнительное исследование трёх алгоритмов подбора весового коэффициента для взвешенного сложения многоканальных АСМ-изображений. Предложенный метод, исключаяющий этап сегментации и анализа гистограмм размеров областей за счёт прямого поиска оптимума от центра диапазона весов, обеспечивает ускорение вычислений в 6 раз относительно эталонного глобального перебора. Отклонения найденных весовых коэффициентов и значения локальной корреляционной метрики статистически незначимо отличаются от результатов сегментного подхода, что подтверждает сохранение точности при

упрощении вычислительного процесса. Гладкий характер зависимостей метрики от веса и наличие единственного максимума обосновывают устойчивость центрированной стратегии поиска. Алгоритм демонстрирует высокую практическую применимость и готов к внедрению в системы автоматизированной обработки многоканальных данных атомно-силовой микроскопии.

COMPARISON OF WEIGHT COEFFICIENT SELECTION ALGORITHMS FOR FUSION OF ATOMIC FORCE MICROSCOPY IMAGES

M.YU LAVETSKI

Abstract. This study addresses the acceleration of weight coefficient selection for weighted fusion of multichannel atomic force microscopy images using a local correlation metric. A segmentation-free algorithm is proposed, which initiates direct optimization from the center of the weight range, achieving a 6.1-fold speedup over exhaustive search on 385 image pairs with statistically insignificant weight deviation.

Keywords: atomic force microscopy, weighted channel combination, local correlation metric, accelerated image processing, Otsu binarization.

Список литературы

1. Jifeng S., Yuanjiao J., Shaoyong Z. // Proceedings of the SPIE International Conference on Space Information Technology. 2008. Vol. S98S. P. 739-744.
2. Zhang A.K., Dare Y.P. // ISPRS Journal of Photogrammetric and Remote Sensing. 2007. Vol.62, No. 4. P.249-263.
3. Ловецкий М.Ю., Цветков В.Ю., Борискевич А.А., Лапицкая В.А., Чижик С.А. // Доклады БГУИР. 2023. Т. 21, № 3. С. 78-86.
4. Petrovic V., Xydeas C. // Tenth IEEE International Conference on Computer Vision (ICCV '05). 2005. Vol. 1, P. 1866-1871.
5. Piella G., Heijmans H. // Proceedings International Conference on Image Processing (Cat. No.03 CH37429). 2003. P. 111-173.
6. Qu G., Zhang D., Yan P. // Opt. Express. 2001. Vol. 9. P. 184-190.
7. Aslantas V., Bendes E. // AEU - International Journal of Electronics and Communications. 2015. P. 1890-1896.

УДК 004.056.5

РАЗРАБОТКА ПРАВИЛ КОРРЕЛЯЦИИ СОБЫТИЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В SIEM-СИСТЕМАХ НА ОСНОВЕ НОРМАТИВНОГО ПЕРЕЧНЯ СОБЫТИЙ

Г.С. СМОТРУК^{1,2}, С.Н. ПЕТРОВ¹, Т.А. ПУЛКО¹

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь,
2 – ООО «СофтЛайн Директ»

Поступила в редакцию 7 апреля 2026

Аннотация. Рассмотрены вопросы повышения эффективности мониторинга информационной безопасности организаций за счет разработки и применения адаптированных правил корреляции событий в системах класса SIEM. Проведен анализ нормативно установленного перечня типов и записей событий информационной безопасности, подлежащих регистрации, консолидации и хранению. Выполнено сопоставление указанных типов событий с конкретными записями журналов логирования различных компонентов информационной инфраструктуры. Разработан набор правил корреляции, ориентированных на выявление типовых угроз и инцидентов информационной безопасности. Результаты апробации разработанных правил корреляции показали снижение уровня ложноположительных срабатываний и повышение эффективности выявления нарушений информационной безопасности.

Ключевые слова: информационная безопасность, SIEM-система, корреляция событий, журналы событий, мониторинг информационной безопасности, киберинциденты, приказ ОАЦ от 25 июля 2023 г. № 130.

Введение

Современные организации функционируют в условиях активной цифровизации и широкого внедрения информационных технологий, что приводит к росту запроса на безопасность. Использование распределенных вычислительных сред, удаленного доступа и сетевых приложений увеличивает сложность информационной инфраструктуры и расширяет поверхность потенциальных атак. Одновременно наблюдается рост числа и сложности кибератак, применяющих многоэтапные сценарии проникновения и маскировку вредоносной активности под легитимные действия пользователей. Одним из основных источников информации о состоянии информационной безопасности являются журналы событий, формируемые различными компонентами информационной инфраструктуры: операционными системами, серверными приложениями, СУБД, сетевым оборудованием и средствами защиты информации. Однако анализ таких журналов по отдельности затруднен из-за большого объема данных, различий в форматах записей и отсутствия единого контекста событий. Для решения данной проблемы применяются системы класса SIEM (Security Information and Event Management), обеспечивающие централизованный сбор, нормализацию, хранение и анализ событий информационной безопасности. Одним из ключевых механизмов анализа событий в таких системах являются правила корреляции, позволяющие выявлять взаимосвязанные события и формировать уведомления о потенциальных инцидентах безопасности.

Актуальность исследования обусловлена также развитием нормативно-правового регулирования в сфере информационной безопасности Республики Беларусь. В частности, в результате принятия приказа Оперативно-аналитического центра при Президенте Республики Беларусь от 25 июля 2023 г. № 130 «О мерах по реализации Указа Президента Республики Беларусь от 14 февраля 2023 г. № 40» (далее – Приказ) установлен обязательный перечень типов

и записей событий информационной безопасности, подлежащих регистрации, консолидации и хранению. Данный перечень определяет минимально необходимый набор событий, подлежащих сбору, и служит основой для построения процессов мониторинга, расследования и реагирования на инциденты информационной безопасности с использованием систем класса SIEM. Перечень служит исходной точкой для последующего сопоставления событий с конкретными источниками логирования и их идентификаторами, а также для разработки правил корреляции, направленных на выявление инцидентов информационной безопасности в SIEM-системах [1].

Несмотря на широкое распространение SIEM систем и наличие в них встроенных наборов правил корреляции, большинство таких правил ориентировано на типовые сценарии угроз и не учитывает особенности конкретной информационной инфраструктуры, а также требования национального законодательства. В результате использование стандартных правил корреляции может либо не обеспечивать достаточный уровень обнаружения инцидентов, либо приводить к значительному количеству ложноположительных срабатываний.

Целью исследования является разработка набора правил корреляции событий информационной безопасности для SIEM-системы (на примере платформы RuSIEM), обеспечивающих автоматизированное выявление значимых событий и инцидентов информационной безопасности на основе нормативно установленного перечня событий.

Сопоставление типов событий с журналами логирования и их идентификаторами

Для выполнения требований Приказа в организации реализована архитектура централизованного сбора и обработки событий информационной безопасности, учитывающая разнородность источников, их критичность и масштаб инфраструктуры. Практика показывает, что нормативно установленный перечень событий может быть собран только при сочетании встроенных механизмов журналирования операционных систем с расширенными средствами аудита, а также при использовании устойчивых каналов доставки и буферизации.

В инфраструктуре на рабочих станциях и части серверов семейства Windows базовым источником является журнал безопасности и системные журналы. Однако для значимой части задач мониторинга встроенного аудита недостаточно, поскольку он фиксирует факт события, но часто не содержит необходимых атрибутов (например, хеши исполняемых файлов, сведения о родительском процессе, детализированную сетевую активность процесса). Поэтому сбор событий в Windows строится в два слоя. Первый слой – расширенный аудит безопасности, включаемый и настраиваемый централизованно через групповые политики домена Active Directory. Он позволяет детализировать аудит процессов, прав доступа, задач планировщика, служб, изменений политик и других категорий. Настройка производится через Advanced Audit Policy Configuration, где включаются категории Process Creation, Logon/Logoff, Account Management, Policy Change, Object Access, Privilege Use и другие. Существенным моментом является включение опции отображения командной строки в событиях создания процессов (Audit Process Creation: Include command line in process creation events), что позволяет для события 4688 получить аргументы запуска и повысить детектируемость скриптовых атак, «живущих за счет встроенных средств» (например, PowerShell, wscript, rundll32) [2].

Второй слой – расширенный журнал событий Sysmon, устанавливаемый как служба и ведущий логирование в отдельный журнал Microsoft-Windows-Sysmon/Operational. Sysmon обеспечивает глубоко контекстные события, которые повышают качество последующей корреляции. Для запуска процессов фиксируются события Sysmon Event ID 1 (Process Create), содержащие командную строку, родительский процесс, путь к бинарному файлу, идентификаторы процесса и пользователя, а при соответствующей конфигурации – хеши. Сетевые соединения процессов регистрируются через Sysmon Event ID 3 (Network Connection), что позволяет связывать сетевую активность не только с хостом, но и с конкретным процессом-инициатором. Для контроля изменений в файловой системе применяется Sysmon Event ID 11 (FileCreate), а изменения в реестре, связанные с закреплением и модификацией конфигурации, фиксируются событиями Sysmon Event ID 12/13/14 (создание ключа, установка значения, переименование значения). Для выявления техник внедрения в процессы и скрытого выполнения полезны Sysmon Event ID 8 (CreateRemoteThread) и Sysmon Event ID 10 (Process Access). Для наблюдения за загрузкой драйверов используется Sysmon Event ID 6 (Driver Load), что актуально

при попытках обхода средств защиты и внедрения вредоносных драйверов. Конфигурация Sysmon определяется XML-политикой и может быть унифицирована на уровне организации: в ней задаются правила include/exclude для событий с фильтрацией по путям, образам процессов, командным строкам, сетевым назначениям и т.п. В результате Windows-источник обогащается деталями (базовые идентификаторы событий из Security/System покрывают нормативную часть, а Sysmon повышает наблюдаемость и точность детектирования на этапе корреляции).

Для операционных систем семейства Linux в качестве базовых источников используются журналы systemd-journal (journal), сообщения подсистемы syslog, а также механизм аудита ядра auditd. Journald и syslog закрывают значительную часть системных событий (запуск/остановка сервисов, сообщения ядра о подключении устройств, сетевые события, сообщения приложений), однако ключевой компонент для выполнения требований по контролю действий пользователей и процессов – auditd. Auditd работает на уровне ядра, фиксируя системные вызовы и формируя события аудита, которые могут использоваться как для расследований, так и для высокоточной корреляции. Практическое преимущество auditd состоит в том, что он позволяет описывать правила фиксации действий независимо от прикладного уровня, то есть регистрировать не «лог приложения», а факт обращения процесса к объекту или вызова системного интерфейса [3].

Правила auditd могут задаваться как в виде набора директив в файлах /etc/audit/rules.d/.rules с последующей компиляцией через augenrules, так и напрямую через auditctl. В правилах определяются системные вызовы (например, execve, openat, chmod, chown, setuid), фильтры по архитектуре (b64/b32), по пользователю (uid/auid), по успешности выполнения и по путям к объектам (watch rules). Для контроля запуска процессов применяется правило аудита системного вызова execve, позволяющее фиксировать выполнение бинарных файлов и параметров запуска с привязкой к пользователю и терминалу. Типичным примером является правило вида -a always, exit -F arch=b64 -S execve -k exec_log, где метка (key) используется для последующей фильтрации и маршрутизации событий при отправке в систему мониторинга. Для контроля изменений критичных файлов (например, /etc/passwd, /etc/shadow, /etc/sudoers, конфигураций сетевых сервисов) используются watch-правила: -w /etc/passwd -p wa -k identity, фиксирующие операции записи и изменения атрибутов. Для контроля действий планировщика могут быть выделены правила на файлы /etc/crontab, каталоги /etc/cron и пользовательские crontab-файлы. Для отслеживания применения привилегий важны события, связанные с sudo (как журналы /var/log/auth.log или /var/log/secure, так и аудит системных вызовов setuid/setgid), а также системные события изменения ролей и контекстов безопасности, если используется SELinux (avc: denied) или AppArmor [9].

Auditd формирует события нескольких типов, наиболее часто используемых для задач мониторинга: EXECVE и SYSCALL для запуска процессов и системных вызовов, USER_LOGIN/USER_LOGOUT/ USER_AUTH для событий аутентификации, USER_ADD/USER_DEL/ USER_MOD для управления учетными записями, SERVICE_START/SERVICE_STOP для жизненного цикла сервисов. В дополнение к auditd важны сообщения ядра и systemd о подключении устройств (USB и другая периферия), а также сетевые события подсистемы фильтрации пакетов. В зависимости от реализации это могут быть сообщения iptables, nftables, firewalld и записи подсистемы netfilter, которые используются для фиксации разрешенных/заблокированных соединений и попыток доступа к нестандартным портам. Для обеспечения непрерывности логирования в Linux применяется ротация журналов (logrotate), управление размером и временем хранения journald, а также буферизация на уровне транспортных сервисов (rsyslog/syslog-ng) при временной недоступности центрального коллектора.

Разработка правил корреляции

В соответствии с изложенными принципами в рамках практической реализации разработан перечень правил корреляции событий информационной безопасности, реализующих требования нормативно установленного набора типов и записей событий. Правила разработаны для применения в SIEM-системе RuSIEM и ориентированы на выявление значимых аномалий и инцидентов информационной безопасности в условиях функционирования реальной информационной инфраструктуры организации. Представленные правила охватывают

ключевые категории источников событий и классы информационных активов, включая операционные системы, системы управления базами данных, телекоммуникационное оборудование, средства защиты информации и прикладное программное обеспечение. Каждое правило построено на анализе формализованных полей нормализованных событий и использует логические, строковые и агрегирующие механизмы корреляции, поддерживаемые SIEM-системой. Список разработанных правил включает правила:

- выявления незапланированного отключения операционной системы;
- контроля событий остановки и аварийного завершения системных служб и демонов;
- обнаружения фактов использования недоверенных съемных машинных носителей информации;
- выявления фактов подключения запрещенных периферийных устройств и каналов передачи данных;
- детектирования попыток установки и запуска недоверенного программного обеспечения с признаками обхода ограничений привилегий;
- обнаружения признаков компрометации учетных записей пользователей;
- детектирования аномального использования привилегированных учетных записей пользователей;
- контроля операций создания, удаления и модификации учетных записей пользователей;
- выявления неудавшихся и отмененных действий пользователей и процессов;
- контроля создания и изменения заданий автозапуска и планировщика задач;
- контроля установки, удаления, перезапуска и ошибок запуска служб и сервисов;
- детектирования изменений системной конфигурации, сетевых настроек и средств межсетевого экранирования;
- обнаружения фактов удаления журналов и следов аудита в операционных системах;
- контроля изменений и попыток изменения настроек средств защиты операционных систем;
- выявления несанкционированных сетевых соединений и удаленного доступа;
- детектирования массового архивирования и вывода данных в операционных системах;
- контроля сессий и аутентификации пользователей в системах управления базами данных;
- обнаружения попыток аутентификации в СУБД с использованием незарегистрированных учетных записей;
- детектирования подозрительных административных действий пользователей в системах управления базами данных;
- контроля операций назначения и отзыва привилегий пользователей в системах управления базами данных;
- контроля запуска и остановки телекоммуникационного оборудования;
- контроля изменений системной конфигурации телекоммуникационного оборудования;
- выявления подозрительных операций с локальными учетными записями на сетевых устройствах;
- обнаружения фактов удаления журналов и следов аудита в операционных системах;
- контроля изменений и попыток изменения настроек средств защиты операционных систем.

Ниже продемонстрировано правило выявления фактов подключения и отключения устройств ввода/вывода к сетевым устройствам. Подключение USB-устройств, модемов и накопителей к сетевым устройствам может использоваться для обхода политики безопасности, загрузки конфигураций или вывода данных (достаточно распространенный случай). Корреляция фиксирует все операции подключения и отключения устройств ввода/вывода.

Huawei: %DEV/5/DEV_ADD, %DEV/5/DEV_REMOVE.

Check Point: USB device inserted/removed.

FortiGate: logid=0100040000.

Cisco: syslog USB insert/remove.

Логика работы правила корреляции:

```
(  
  device.vendor in {"Huawei","Check Point","Fortinet","Cisco"}  
)
```

AND
 (
 event.type contains "DEV_ADD"
 OR event.type contains "DEV_REMOVE"
 OR event.type contains "USB"

Обобщенная модель корреляции событий информационной безопасности представлена в табл. 1.

Табл. 1. Обобщенная модель корреляции событий информационной безопасности

Категория события	Источник	Признаки события	Условия корреляции	Выявляемый инцидент
Аутентификация пользователей	Windows Security Log / Linux auth.log	Множественные ошибки входа	>5 неудачных попыток входа за короткий интервал времени	Подбор пароля (Brute Force)
Аутентификация пользователей	Windows / Linux	Успешный вход после серии ошибок	Успешная аутентификация после множества неудачных попыток	Компрометация учетной записи
Управление учетными записями	Active Directory / Linux user management	Создание новой учетной записи	Создание пользователя вне рабочего времени	Несанкционированное создание учетной записи
Управление учетными записями	Active Directory	Добавление пользователя в группу администраторов	Добавление в привилегированную группу	Эскалация привилегий
Использование привилегий	Windows Security Log / Linux sudo	Использование административных прав	Частые вызовы привилегированных команд	Подозрительная административная активность
Запуск процессов	Windows Process Events / Linux auditd	Запуск системных утилит	Запуск инструментов администрирования обычным пользователем	Возможная подготовка атаки
Изменение системных настроек	Windows Security Policy / Linux configuration logs	Изменение политики безопасности	Изменение параметров аудита или политики безопасности	Попытка сокрытия активности
Доступ к файлам	Windows Object Access / Linux file audit	Массовое чтение или копирование файлов	Большое количество операций чтения за короткое время	Возможная утечка данных
Сетевые соединения	Firewall / IDS / Netflow	Подключения к подозрительным адресам	Повторяющиеся соединения с неизвестными внешними узлами	Командный сервер (C2)
Сетевые соединения	Firewall logs	Массовые соединения на разные порты	Сканирование портов	Сетевая разведка
Аппаратные события	Windows Device Events / Linux udev	Подключение внешних устройств	Подключение USB в защищенной зоне	Потенциальная утечка данных
Завершение процессов	Windows / Linux	Принудительное завершение системных служб	Остановка служб безопасности	Отключение средств защиты

Проведенный анализ журналов операционных систем, сетевого оборудования и средств защиты информации позволил выделить основные категории событий, имеющих наибольшее значение для задач мониторинга информационной безопасности. К ним относятся события аутентификации пользователей, управление учетными записями, использование привилегий, запуск процессов, доступ к данным и сетевые соединения. Несмотря на различия в форматах журналов и идентификаторах событий различных операционных систем, большинство событий могут быть приведены к унифицированным логическим типам. Это позволяет формировать универсальные правила корреляции, основанные на семантике событий, а не на конкретных

идентификаторах журналов. Разработанная обобщенная модель корреляции связывает категории событий, источники журналов и условия их совместного анализа, что позволяет выявлять типовые сценарии нарушений информационной безопасности, включая подбор паролей, компрометацию учетных записей, эскалацию привилегий, подготовку атак, утечку данных и взаимодействие с командными серверами злоумышленников.

Заключение

Исследование подтвердило важность системного подхода к разработке правил корреляции событий информационной безопасности в SIEM-системах. В отличие от типовых наборов правил, ориентированных на универсальные сценарии угроз, предложенный подход основан на учете нормативных требований и особенностей функционирования конкретной информационной инфраструктуры. Сформирована методика сопоставления нормативно установленных в соответствии с Приказом типов событий информационной безопасности с записями журналов различных источников, что обеспечивает корректную интерпретацию событий на уровне системы мониторинга безопасности. Разработан набор правил корреляции, позволяющий выявлять сложные сценарии атак и нарушений политики безопасности. Апробация разработанных правил показала их высокую эффективность при эксплуатации в реальной информационной инфраструктуре. Применение предложенных правил позволило повысить и снизить количество ложноположительных срабатываний и улучшить процессы реагирования на события информационной безопасности. Полученные результаты могут быть использованы при проектировании и модернизации систем мониторинга безопасности, а также при разработке нормативно ориентированных механизмов обнаружения инцидентов информационной безопасности.

DEVELOPMENT OF RULES FOR CORRELATION OF INFORMATION SECURITY EVENTS IN SIEM SYSTEMS BASED ON A NORMATIVE LIST OF EVENTS

H.S. SMATRUK, S.N. PETROV, T.A. PULKO

Abstract. The issues of increasing the effectiveness of monitoring the information security of organizations through the development and application of adapted rules for event correlation in SIEM class systems are considered. The analysis of the normatively established list of types and records of information security events subject to registration, consolidation and storage is carried out. The specified types of events were compared with specific log entries of various components of the information infrastructure. A set of correlation rules has been developed aimed at identifying typical threats and incidents of information security. The results of testing the developed correlation rules showed a decrease in the level of false positive positives and an increase in the effectiveness of detecting information security violations.

Keywords: information security, SIEM system, event correlation, event logs, information security monitoring, cyber incidents, OAC Order No. 130 dated July 25, 2023.

Список литературы

1. О мерах по реализации Указа Президента Республики Беларусь от 14 февраля 2023 г. № 40 : приказ Оператив.-аналит. центра при Президенте Респ. Беларусь, 25 июля 2023 г., № 130 // Национальный правовой Интернет-портал Республики Беларусь. – URL: <https://pravo.by/document/?guid=12551&p0=T62305425> (дата обращения: 06.04.2026).
2. Таненбаум Э. Современные операционные системы / Э. Таненбаум, Х. Бос. – 4-е изд. – СПб.: Питер, 2015. – 1120 с.
3. Столлингс У. Операционные системы. Внутренняя структура и принципы проектирования / У. Столлингс. – 9-е изд. – М.: Вильямс, 2019. – 944 с.
4. Робачевский А. М. Операционная система UNIX / А. М. Робачевский. – СПб.: БХВ-Петербург, 2016. – 640 с.

КОНЕЧНЫЕ РАЗНОСТИ В ЗАДАЧЕ СИНТЕЗА ВЕРОЯТНОСТИ ПРОПУСКА ОШИБКИ ПРИ НАБЛЮДЕНИИ ВЕКТОРОВ ПЕРЕХОДОВ

И.П. КОБЯК

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 5 апреля 2026

Аннотация. В работе рассмотрена методика формирования фрагмента производящей функции для вероятности пропуска ошибки предлагаемым методом свертки. При этом использованы частные реализации функции правдоподобия. Для достижения поставленной цели применена методика синтеза конечных разностей с последующей подстановкой результатов в исходное равенство. Основное назначение полученных равенств заключается в возможности определения численного значения интегральной площади под кривой плотности распределения вероятностей пропуска ошибки при наблюдении векторов переходов. Заданные системой наблюдения векторы рассматриваются в данной задаче в виде подмножества сложных событий в множестве всех векторов переходов.

Ключевые слова: производящая функция, сложные события, векторы переходов, вероятность пропуска ошибки, интеграл степенного ряда, факториальные моменты.

Введение

Как правило, в основной массе передаваемых сообщений используются идентификаторы на основе выборочных функций [1] или линейной свертки [2], характеризующих свойства сообщений конечной длины. Данные подходы определяют необходимость анализа алгоритмов синтеза контрольной свертки с учетом ряда научно-технических направлений. В частности, при конкретной реализации сообщений необходимо знать уровень коллизий (или обнаруживающие свойства) точечных оценок или линейной сигнатуры. Это позволяет сократить время на поиск и исправление возможных ошибок в случае непредвиденных искажений данных. С общетеоретической точки зрения необходим системный анализ свойств оценок на основе производящих функций (ПФ). Результатом таких исследований с учетом принципов формализации зависимостей является возможность выбора наиболее эффективного из алгоритмов контрольного кодирования путем сопоставления свойств кодов по ряду выбранных показателей [1]. В некоторых случаях при абонентском обмене данными на фоне шума необходимо знать теоретические уровни вероятностных параметров таких как математическое ожидание, мода и площадь под интегральной кривой плотности распределения вероятностей ошибки. Однако, применение новых методов идентификации (в части обоснования целесообразности) как правило требует весьма сложных вычислений для формирования численных значений производящих функций.

В связи с указанными обстоятельствами в данной работе рассматривается метод синтеза точечных оценок вероятности наблюдения векторов переходов заданного вида (ВПЗВ) с целью последующего интегрального сравнения свойств представляемой (частной) свертки с общеизвестными методами идентификации. Однако множество промежуточных вычислений, необходимых для решения всей задачи синтеза ПФ, приводит к возможности представления в данной статье только частных моментов решения данной проблемы. При этом число ВПЗВ в данной задаче равно $j = 1$.

Разности для нижней границы вероятности пропуска ошибки

Нижняя граница для интеграла плотности распределения вероятностей пропуска ошибки для всех значений разрядности векторов $r > 1$ определяется максимальной величиной параметрической ортонормированной вероятности $\frac{3^{r-\mu}}{2m^2}$ при $r = 2$, что следует из тривиальных подстановок параметров r и μ в приведенную формулу. Здесь μ это число бит перехода вида 01 в регистрируемой паре векторов.

Рассмотрим теперь разность двух объектов с учетом результатов [3], определяющих функцию правдоподобия для числа векторов переходов с $j = 1$ и произвольных значений i

$$\frac{1}{2^{i+1}}\beta_i - \frac{1}{2^{i+2}}\beta_{i+1} = \frac{1}{2^{i+1}}\beta_{i+1} \frac{1}{\sqrt{1-4p}} \left[\left(1 + \sqrt{1-4p}\right)^{i+1} - \left(1 - \sqrt{1-4p}\right)^{i+1} \right] - \\ - \frac{1}{2^{i+2}} \frac{1}{\sqrt{1-4p}} \left[\left(1 + \sqrt{1-4p}\right)^{i+2} - \left(1 - \sqrt{1-4p}\right)^{i+2} \right]$$

Соответственно при заданном выше уровне вероятности p достаточно просто получить соотношение

$$\frac{1}{2^{i+2}}\beta_{i+1} = \frac{1}{2^{i+1}}\beta_i - \sqrt{\frac{9}{160}} \frac{1}{2^{i+1}} \left[\left(1 + \sqrt{\frac{5}{8}}\right)^i - \left(1 - \sqrt{\frac{5}{8}}\right)^i \right].$$

Получим частные реализации данной разности, выразив все значения для $i = 3, 4, 5$ через элементы $\frac{1}{2^{i+1}}\beta_i$, $i = 2, 3, 4$, тогда

$$\frac{1}{2^4}\beta_3 = \frac{1}{2^3}\beta_2 - \sqrt{\frac{9}{160}} \frac{1}{2^3} \left[\left(1 + \sqrt{\frac{5}{8}}\right)^2 - \left(1 - \sqrt{\frac{5}{8}}\right)^2 \right],$$

$$\frac{1}{2^5}\beta_4 = \frac{1}{2^3}\beta_2 - \sqrt{\frac{9}{160}} \sum_{l=2}^3 \frac{1}{2^{l+1}} \left[\left(1 + \sqrt{\frac{5}{8}}\right)^l - \left(1 - \sqrt{\frac{5}{8}}\right)^l \right],$$

$$\frac{1}{2^6}\beta_5 = \frac{1}{2^3}\beta_2 - \sqrt{\frac{9}{160}} \sum_{l=2}^4 \frac{1}{2^{l+1}} \left[\left(1 + \sqrt{\frac{5}{8}}\right)^l - \left(1 - \sqrt{\frac{5}{8}}\right)^l \right].$$

Для граничного значения параметра i имеем

$$\frac{1}{2^{n-1}}\beta_{n-3} = \frac{1}{2^3}\beta_2 - \sqrt{\frac{9}{160}} \sum_{l=2}^{n-3} \frac{1}{2^{l+1}} \left[\left(1 + \sqrt{\frac{5}{8}}\right)^l - \left(1 - \sqrt{\frac{5}{8}}\right)^l \right].$$

Данные равенства предполагают суммирование рядов бесконечной выборки с учетом приведенных составляющих производящей функции.

Суммирование факториальных моментов в ПФ

Для первой суммы моментов в ПФ, соответствующей первой степени вероятности p запишем равенство вида

$$PF_1 = Mo + \left[\xi_1 p \left(1 + \frac{1}{2^3} \beta_2 + \frac{1}{2^4} \beta_3 + \dots + \frac{1}{2^{n-2}} \beta_{n-3} + \frac{1}{2^{n-1}} \beta_{n-2} \right) \right]^{0,5n-np},$$

где значение $\xi_1 = 1 - p$, что следует из квадратичной формы для параметрической вероятности.

Суммирование факториальных моментов в скобках позволяет сформировать равенство

$$\begin{aligned} & \frac{1}{2^3} \beta_2 + \frac{1}{2^4} \beta_3 + \dots + \frac{1}{2^{n-2}} \beta_{n-3} + \frac{1}{2^{n-1}} \beta_{n-2} = \\ & = (n-3) \frac{1}{2^3} \beta_2 - \sqrt{\frac{9}{160}} \sum_{l=2}^{n-3} (n-2-i) \frac{1}{2^{i+1}} \left[\left(1 + \sqrt{\frac{5}{8}} \right)^i - \left(1 - \sqrt{\frac{5}{8}} \right)^i \right] + \\ & + \frac{1}{2^{n-2}} - \sqrt{\frac{8}{5}} \left[\left(1 + \sqrt{\frac{5}{8}} \right)^{n-3} - \left(1 - \sqrt{\frac{5}{8}} \right)^{n-3} \right]. \end{aligned}$$

Вычислим вычитаемый элемент данного ряда преобразуя его к виду

$$(n-2) \frac{1}{2} \sqrt{\frac{9}{160}} \sum_{l=2}^{n-3} \left[\left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i - \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i \right] - \frac{1}{2} \sqrt{\frac{9}{160}} \sum_{l=2}^{n-3} i \left[\left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i - \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i \right].$$

При этом для упрощения расчетов введем сокращение вида

$$(n-2) \frac{1}{2} \sqrt{\frac{9}{160}} \sum_{l=2}^{n-3} \left[\left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i - \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i \right] = (n-2) \left(\frac{261}{1024} + \frac{1}{2} \sqrt{\frac{9}{160}} \sum_{l=5}^{n-3} a^i \right).$$

В данном соотношении учтено, что при $i = 2, 3, 4$ точное значение вычисляемого параметра равно $\frac{261}{1024}$. Таким образом, нижний предел суммирования оказывается равным пяти, а вычитаемый элемент

$$\left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{5}{8}} \right)^5 = 1,259 \cdot 10^{-5}$$

для данной и более высоких степеней не учитывается.

Теперь, используя принцип суммирования арифметических прогрессий, можем записать

$$i = 2, 3, \dots, n-3, \quad (a^2 + a^3 + a^4) + a^5 + \dots + a^{n-3} + a^{n-2}$$

Соответственно получаем:

$$(n-2) \left[\frac{261}{1024} + \frac{1}{2} \sqrt{\frac{9}{160}} \left(a^2 \frac{1}{1-a} - a^2 \frac{1-a^3}{1-a} \right) \right] = (n-2) \left(\frac{261}{1024} + \frac{667}{1024} \right) = \frac{928}{1024} (n-2). \quad (1)$$

Далее будем учитывать, что данное значение соответствует равенству $(1-p)(n-2)$.

Суммирование рядов вычитаемых моментов в функции правдоподобия.

Второй элемент вычитаемого в сумме моментов представляется со знаком плюс и имеет вид

$$f(i) = \frac{1}{2} \sqrt{\frac{9}{160}} i \left[\left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i - \left(\frac{1}{2} - \frac{1}{2} \sqrt{\frac{5}{8}} \right)^i \right].$$

При этом для трех первых элементов суммы имеем значение $\frac{1077}{1024}$. Остальные слагаемые вычисляем по упрощенной формуле

$$f(i) = \frac{1077}{1024} + \frac{1}{2} \sqrt{\frac{9}{160}} \sum_{i=5}^{n-3} i a^i.$$

Данное соотношение позволяет записать равенство с учетом прогрессии

$$f(i) = \frac{1077}{1024} + \frac{1}{2} \sqrt{\frac{9}{160}} a \left[5a^4 + 6a^5 + \dots + (n-3)a^{n-4} \right].$$

Соответственно, используя механизм интегрирования рядов и дифференцирования суммы, имеем

$$a \left[5 \int a^4 + 6 \int a^5 + \dots + (n-3) \int a^{n-4} \right] = a (a^5 + a^6 + \dots + a^{n-3}) = \frac{a^6}{1-a}.$$

$$\frac{\partial}{\partial a} \frac{a^6}{1-a} = 6a^5 \frac{1}{1-a} + a^6 (1-a)^{-2}. \quad (2)$$

Численное значение производящей функции с учетом приведенных выше результатов (1) и (2) может быть определено по формуле

$$PF_1 = Mo + \left[\xi_1 p \left(1 + \frac{29}{32} (n-3) - \frac{928}{1024} (n-2) + \frac{1077}{1024} + \frac{1}{2} \sqrt{\frac{9}{160}} \left(6a^5 \frac{1}{1-a} + a^6 (1-a)^{-2} \right) \right]^{0,5n-pn} = Mo + (0,966)^n ..$$

В приведенных вычислениях используется несуществующий момент для $i = n - 3$ что не вносит практически коррективов в полученные результаты.

Заключение

В представленной работе рассмотрены конечные разности факториальных моментов, характеризующих параметры оценок с теоретической точки зрения. Используемые алгоритмы преобразования сложных рядов позволили сформировать частный случай производящей функции, соответствующей наблюдению векторов переходов при $j = 1$.

Полученный результат, с учетом одновременного не совпадения в сумме моды и значения параметра в скобках, свидетельствует о высоких вероятностных показателях оценок при $n \gg 1$ вероятности наблюдения ВПЗВ в каналах связи.

FINITE DIFFERENCES IN THE PROBABILITY SYNTHESIS PROBLEM OF MISSING ERRORS WHEN OBSERVING VECTORS

I.P. KOBIAK

Abstract. The paper considers a technique for forming a fragment of a generating function for the probability of missing an error by the proposed new convolution method. To achieve this goal, a technique for synthesizing finite differences was applied, followed by substituting the results into

the initial equality. The main purpose of the obtained equalities is to be able to determine the numerical value of the integral area under the density curve of the probability distribution of error omission when observing transition vectors. The vectors specified by the observation system are considered in this problem as a subset of specified complex events in the set of all transition vectors.

Keywords: generating function, complex events, transition vectors, probability of missing an error, power series integral, factorial moments.

Список литературы

1. Kobyak I.P. On the boundaries of probabilistic arguments in the synthesis of linear signatures and statistical estimates. *Izvestiya vuzov. Electronics*. 2020. T. 25. No. 2. Pp. 175-182.
2. Gordon G., Nadig H. Fault localization in microprocessor systems using hexadecimal key codes. *Electronics*. 1977. No. 5. Pp. 89-96
3. Riordan J. Combinatorial identities. M. Nauka. 1982.

МОДЕЛЬ ОПРЕДЕЛЕНИЯ ДИСТАНЦИИ ДО ОБЪЕКТА В КВАЗИСТАЦИОНАРНОЙ МНОГОРОБОТНОЙ СИСТЕМЕ БЕЗ ОДОМЕТРИИ НА ОСНОВЕ СОГЛАСОВАННОЙ ТРИАНГУЛЯЦИИ

Ц. ГО¹, М.И. ЗОРЬКО¹, С.В. ЛИТВИНОВ², В.Ю. ЦВЕТКОВ¹

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

2 – Холдинг «АБВ», Российская Федерация

Поступила в редакцию 12 апреля 2026

Аннотация. Рассматриваются модели определения дистанции до объекта окружения одним роботом и в децентрализованной самоорганизующейся видео-центрической многороботной системе многороботной системе без одометрии (POM-MRS), функционирующей в квазистационарном режиме. Для повышения точности определения дистанции до объекта окружения в POM-MRS разработана модель на основе локальной согласованной триангуляции с аналитической оценкой ошибок измерений и протоколно-геометрической верификацией гипотез о детектировании агентов, описывающая базовые функции, позволяющие построить общий алгоритм определения дистанции: определение распределенного стереобазиса; вычисление дистанции до объекта окружения; локальная согласованная триангуляция, включающая выбор опорного агента (статическая оптимизация локального распределенного стереобазиса) и управляемое перемещение опорного агента (динамическая оптимизация распределенного стереобазиса); локальная визуально-протоколная идентификация агентов на основе протоколно-геометрической верификации гипотез об их детектировании. Показано, что предложенная модель, позволяет уменьшить ошибки определения дистанции до объекта окружения по сравнению с однороботными моделями на основе одометрии различного вида и известными моделями POM-MRS с комплексным нейросетевым детектированием агентов.

Ключевые слова: определение дистанции до объекта, видео-центрическая самоорганизующаяся многороботная система без одометрии.

Введение

Повышение точности определения дистанции до объекта имеет критически важное значения для построения эффективных систем локализации и картирования SLAM (Simultaneous Localization and Mapping) [1–4]. Преобразуя сигналы сенсоров в понимание окружающего пространства технология SLAM позволяет мобильному роботу ориентироваться в пространстве. Самоорганизующаяся многороботная система (SMRS – Self-Organizing Multi-Robot System) опирается на технологию MA-SLAM (Multi-Agent SLAM), позволяющую агентам совместно оценивать и синхронизировать понимание окружения [4–7]. Дистанция до объектов окружения используется для: вычисления координат и регистрации на распределенной карте объекта с использованием известных координат и ориентаций агентов или измеренного с помощью одометрии перемещения одиночного робота (задача картирования); вычисления координат и ориентации агента или одиночного робота с использованием координат наблюдаемого объекта, взятых из распределенной карты после его идентификации с объектом карты (задача локализации).

SMRS на основе малоразмерных агентов имеют ограниченные возможности применения габаритных высокоточных комбинированных сенсоров: высокоточные лидары и промышленные стереокамеры имеют значительный вес и объем (маленький агент не обладает достаточной грузоподъемностью и полезной площадью для их размещения); «умные» сенсоры с активным

излучением и встроенными вычислителями для обработки изображений потребляют много энергии (для малоразмерного агента с небольшой аккумуляторной батареей это критично снижает время автономной работы). В [7–11] описаны варианты видео-центрических SMRS (V-SMRS), в которых агенты оснащаются одной монокулярной камерой. В V-SMRS у агента отсутствует прямая информация о глубине окружения (расстояниях до объектов) – проблема монокулярного SLAM и визуальной одометрии. Однако, технология MA-SLAM позволяет не только компенсировать недостатки монокулярной камеры, но и получить ряд преимуществ, связанных с динамическим распределенным стереобазисом (регулирование стереобазиса между агентами для повышения точности триангуляции); устранением неопределенности масштаба (стереобазис между агентами может быть определен с помощью монокулярной видеокамеры по известному размеру объекта); согласованной триангуляцией (наблюдение объекта под разными углами множеством агентов позволяет уменьшить ошибку определения дистанции до объекта за счет статистической обработки с использованием фильтра Калмана). Основной проблемой V-SMRS является синхронизация распределенных измерений.

Ранние методы организации V-SMRS использовали, как правило, централизованную модель локализации агентов и формирования карты. Например, предложенный в [6] метод использует объекты окружения БЛА в качестве ориентиров, решая проблему масштаба и перехода к метрическим размерам с помощью барометрического высотомера и инерциальной системы, а проблему определения взаимного положения агентов – с помощью сервера на основе сопоставления общих 3D-точек окружения.

В отдельный класс выделяются V-SMRS без одометрии – POM-MRS (Pure Optical Monocular Multi-Robot System), агенты которых оснащены только монокулярными камерами и, поэтому, могут иметь компактное исполнение. В современных концепциях POM-MRS преобладают децентрализованная модель и комплексное нейросетевое детектирование агентов (включает обнаружение, трекинг, сетевую идентификацию, распознавание относительной ориентации) при определении распределенного стереобазиса с помощью метода обратной проекции (монокулярной дальнометрии) по их известным линейным размерам или известным размерам маркеров на корпусах агентов [12]. Для идентификации агентов в таких системах (установления соответствия между изображениями агентов и их сетевыми адресами) используются маркеры и индикаторы на корпусах агентов, а также совместно наблюдаемые объекты окружения (через функции MA-SLAM) [4]. Комплексное нейросетевое детектирование агентов позволяет существенно повысить точность определения стереобазиса, но на малых дистанциях (до 7 м при использовании широкоугольной оптики), что не всегда согласуется с дистанциями до объектов окружения и может приводить к большим ошибкам их измерения. Использование для идентификации маркеров и объектов окружения также ограничивает расстояние между агентами. Индикаторы демаскируют агентов и не могут использоваться в некоторых приложениях.

Целью работы является определение ограничений использования моделей определения дистанции до объекта окружения в однороботных и многороботных системах.

Модели определения дистанции до объекта окружения в однороботных системах

Для определения дистанции до неподвижных объектов окружения в однороботных системах (SRS – Single Robot System) достаточно эффективна триангуляция. В SRS дистанция $D_O(t_1)$ от робота до объекта определяется (оценка $D_O(t_1)$ или $D_O(t_2)$) с ошибкой $E_D(t_1) = |D_O(t_1) - D_O(t_1)|$ с помощью триангуляции и сопоставления двух перекрывающихся изображений – стереопаре $\{I_O(t_1), I_O(t_2)\}$ (рис. 1), полученных в результате перемещения $B(t_1, t_2) = |\overline{C_{ZYX}}(t_1) - \overline{C_{ZYX}}(t_2)|$ робота между пространственными координатами $\overline{C_{ZYX}}(t_1)$ и $\overline{C_{ZYX}}(t_2)$ в моменты времени t_1 и t_2 , измеряемого (оценка $B(t_1, t_2)$ стереобазиса) с помощью механической, инерциальной, визуальной и других видов одометрии с ошибкой $E_B(t_1, t_2) = |B(t_1, t_2) - B(t_1, t_2)|$, зависящей от качества измерительной системы и сложности условий измерения [13]

$$D_O(t_1) = B(t_1, t_2) \frac{\sin(\alpha_O(t_2))}{\sin(\alpha_O(t_1) + \alpha_O(t_2))}, \quad (1)$$

где $\alpha_O(t_1)$, $\alpha_O(t_2)$ – оценки углов на объект для положений робота в моменты времени t_1 и t_2 , определяемые по изображениям $\{I_O(t_1), I_O(t_2)\}$ с ошибками $E_\alpha(t) = |\alpha_O(t) - \alpha_O(t)|$ относительно истинных значений углов на объект $\alpha_O(t_1)$, $\alpha_O(t_2)$, которые зависят от $B(t_1, t_2)$: $\alpha_O(t_1) = f(B(t_1, t_2))$, $\alpha_O(t_2) = f(B(t_1, t_2))$.

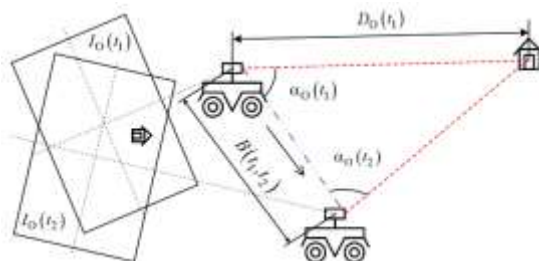


Рис. 1. Схема определения дистанции до объекта в SRS на основе триангуляции

Из выражения (1) следует, что $E_D(t_1) = f(E_B(t_1, t_2), E_\alpha(t_1), E_\alpha(t_2))$. Ошибки $E_B(t_1, t_2)$ определения стереобазиса в однороботных системах растут линейно с увеличением длины $L_S(t_1, t_2)$ пути робота за интервал времени между моментами t_1 и t_2 , т.е. $E_B(t_1, t_2) = f(L_S(t_1, t_2))$. Значения δ относительных погрешностей измерения перемещения с помощью одометрии различного типа для некоторых однороботных систем приведены в таблице 1. [14–20].

Табл. 1. Относительные погрешности определения перемещения робота

R	T_R	T_O	T_S	T_C	T_T	T_A	T_U	δ , %	N
[14]	W	M	P	H	D	m	–	2,0	WM/PHDm0
		V	P	H	D	m	–	1,4	WV/PHDm0
[15]	W	V	O	H	D	w	–	5,2	WV/OHDw0
					L	w	–	3,0	WV/OHLw0
[6, 17]	W	M	P	H	D	w	–	7,0	WM/PHDw0
					D	s	–	0,4	WM/PHDs0
[18]	W	Im / Ime	O	M	D	w	10	8,7 / 1,8	WIm/OMDw10 / WIme/OMDw10
					D	w	15	8,4 / 1,2	WIm/OMDw15 / WIme/OMDw15
					D	w	20	15,8 / 1,2	WIm/OMDw20 / WIme/OMDw20
					D	w	30	31,4 / 0,5	WIm/OMDw30 / WIme/OMDw30
				S	D	w	10	2,9 / 1,2	WIm/OSDw10 / WIme/OSDw10
					D	w	15	11,1 / 1,8	WIm/OSDw15 / WIme/OSDw15
					D	w	20	20,1 / 1,6	WIm/OSDw20 / WIme/OSDw20
					D	w	30	14,0 / 2,7	WIm/OSDw30 / WIme/OSDw30
[19]	Q	V	O	T	D	–	3	1,6	QV/OTD-3
					D	–	5	1,7	QV/OTD-5
[20]	W	V	P	F	D	w	–	53,1	WV/PFDw0
		M	P	F	D	w	–	900	WM/PFDw0
		V	O	M	D	w	–	14,8	WV/OMDw0
		M	O	M	D	w	–	18,5	WM/OMDw0

В табл. использованы следующие обозначения: R – ссылка на источник; T_R – тип робота, $T_R = \{W, Q\}$ (W – колесный робот, Q – квадрокоптер); T_O – тип одометрии, $T_O = \{M, V, Im, Ime\}$ (M – механическая, V – визуальная, Im – комбинированная инерциально-механическая, Ime –

комбинированная электронная инерциально-механическая); T_S – тип окружающего пространства, $T_S = \{P, O\}$ (P – помещение, O – открытое пространство); T_C – качество покрытия (среды) для перемещения робота, $T_C = \{H, M, S, F, T\}$ (H – жесткая, M – средний песок, S – камень и песок, F – мелкий песок, T – турбулентность (для $T_R = Q$)); T_T – тип траектории, $T_T = \{D, L\}$ (D – прямая, L – циклическая); T_A – качество сцепления колес, $T_A = \{w, m, s\}$ (w – слабое, m – среднее, s – сильное); δ – относительная погрешность определения расстояния; N – обозначение варианта комбинации однороботной системы с одометрией определенного типа (параметры $T_R | T_O$) и условий измерения стереобазы (перемещения робота) (параметры $T_S | T_C | T_T | T_A$); T_U – неровности поверхности (см) или высота полета (м) для $T_R = Q$. Системы [14, 15] используют известные расстояния от камеры до поверхности, визуальная одометрия в системе [20] основана на анализе изображения поверхности.

Из табл. следует, что наибольшую точность измерения перемещения обеспечивает механическая одометрия колесных роботов ($\delta \approx 0,4\%$ для $N = WM/PHDs0$). Однако, она очень чувствительна к условиям измерений, особенно к качеству поверхности. При слабом сцеплении колес с поверхностью ошибки механической одометрии резко возрастают ($\delta \approx 18,1..900\%$ для $N = \{WM/OMDw0, WM/PFDw0\}$). Уменьшить погрешности измерений перемещений робота позволяет комбинирование механической и других видов одометрии: инерциально-механическая одометрия ($\delta \approx 2,9..31,4\%$ для $N = \{WIm/OSDw10, WIm/OMDw30\}$) позволяет уменьшить относительную погрешность измерения перемещения колесного робота на песчаном грунте до 310 раз по сравнению с механической одометрией ($\delta \approx 18,1..900\%$ для $N = \{WM/OMDw0, WM/PFDw0\}$); контроль тока потребления двигателя в системе электронной инерциально-механической одометрии ($\delta \approx 0,5..2,7\%$ для $N = \{WIme/OMDw30, WIme/OSDw30\}$) позволяет уменьшить относительную погрешность измерения перемещения колесного робота до 62,8 раза по сравнению с инерциально-механической одометрией ($\delta \approx 2,9..31,4\%$ для $N = \{WIm/OSDw10, WIm/OMDw30\}$). Ошибки визуальной одометрии зависят от условий регистрации изображений и их соответствия условиям калибровки камеры. При хорошей освещенности и высокой точности калибровки камеры визуальная одометрия ($\delta \approx 1,4\%$ для $N = WV/PHDm0$, и $\delta \approx 1,6..1,7\%$ для $N = \{QV/OTD-3, QV/OTD-5\}$) обеспечивает относительные погрешности измерений перемещений робота, соизмеримые с механической ($\delta \approx 2,0\%$ для $N = WM/PHDm0$) и комбинированными видами ($\delta \approx 0,5..31,4\%$ для $N = \{WIme/OMDw30, WIm/OMDw30\}$) одометрии. В открытом пространстве значения относительной погрешности измерения перемещения робота лежат в диапазонах: $\delta \approx 18,1..900\%$ для механической одометрии ($N = \{WM/OMDw0, WM/PFDw0\}$); $\delta \approx 2,9..31,4\%$ для инерциально-механической одометрии ($N = \{WIm/OSDw10, WIm/OMDw30\}$); $\delta \approx 0,5..2,7\%$ для электронной инерциально-механической одометрии ($N = \{WIme/OMDw30, WIme/OSDw30\}$); $\delta \approx 1,6..53,1\%$ для визуальной одометрии ($N = \{QV/OTD-3, WV/PFDw0\}$).

Ошибки $E_a(t)$ зависят не от длины $L_S(t_1, t_2)$ пути робота, а от стереобазиса $B(t_1, t_2)$, т.е. $(E_a(t_1) = f(B(t_1, t_2)), E_a(t_2) = f(B(t_1, t_2)))$ и растут при уменьшении стереобазы из-за приближения углов $\alpha_O(t_1), \alpha_O(t_2)$ к $\pi/2$. Поэтому для минимизации ошибки $E_D(t_1)$ необходимо согласовать $B(t_1, t_2)$ с $D_O(t_1)$ за счет перемещения робота таким образом, чтобы значения углов $\alpha_O(t_1), \alpha_O(t_2)$ попадали на близкие к линейным участки функций $\sin(\alpha_O(t_1))$ и $\sin(\pi - (\sin(\alpha_O(t_1)) + \alpha_O(t_2)))$. Однако данная возможность повышения точности определения расстояния до объекта имеет ограничение, связанное с быстрым ростом ошибок определения значений углов $\alpha_O(t_1), \alpha_O(t_2)$ при одновременном перемещении робота и объекта.

Преодолеть данное ограничение позволяет метод обратной проекции (монокулярной дальнометрии), согласно которому дистанция $D_O(t)$ до объекта известного размера $S_O(A_C)$ вдоль оси A_C определяется по размеру его проекции [12] (рис. 2). В таком случае в любой момент времени t оценка $D_O(t)$ дистанции до объекта вычисляется по оценке $\alpha_O(A_C)$ углового размера $\alpha_O(A_C)$ объекта вдоль предполагаемого положения A_C (определяется по изображению объекта) оси A_C на основе известного углового разрешения $R_M(A_C)$ пикселя фотоприемной матрицы вдоль предполагаемого положения A_C оси A_C с помощью выражения

$$D_O(t) = S_O(A_C) / \tan(\alpha_O(A_C)), \quad (2)$$

где $\alpha_O(A_C) = P_O(A_C) R_M(A_C)$; $R_M(A_C) = \alpha_M(A_C) / P_M(A_C)$; $P_O(A_C)$ – оценка пиксельного размера объекта вдоль предполагаемого положения A_C оси A_C на сформированном фотоприемной матрицей изображении; $\alpha_M(A_C)$, $P_M(A_C)$ – угловой и пиксельный размеры фотоприемной матрицы вдоль предполагаемого положения A_C оси A_C .

Недостатком данного метода определения дистанции является ограниченный набор классов объектов с априорно известными размерами. Однако в POM-MRS размеры агентов известны.

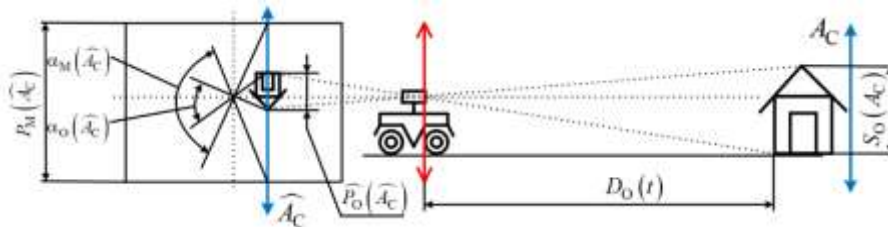


Рис. 2. Схема определения дистанции до объекта в SRS на основе метода обратной проекции

Недостатком данного метода является ограниченный набор классов объектов с априорно известными размерами. Однако в POM-MRS размеры агентов известны.

Модели определения дистанции до объекта окружения в многороботных системах

В случае неподвижных объектов окружения POM-MRS используют модель (2) определения распределенного стереобазиса на основе взаимного наблюдения агентов, а затем модель (1) для определения дистанции до объекта окружения на основе согласованной триангуляции. Для согласованной триангуляции в POM-MRS, как и в SMRS с большим количеством агентов используются алгоритмы консенсуса [21–23], усредняющие результаты измерения дистанции отдельными агентами. Данный подход позволяет достичь высокой точности, но для согласования распределенного стереобазиса с дистанцией до объекта требуется масштабирование пространственной топологии группы агентов или POM-MRS в целом.

Более эффективным подходом с точки зрения экономии ресурсов является локальная согласованная триангуляция на основе двух агентов POM-MRS, временно объединившихся для решения задачи определения дистанции до объекта окружения. В работе [24] обосновывается угол, под которым два агента SMRS должны наблюдать объект, чтобы ошибка определения дистанции была минимальной. Показано, что оптимальным является угол 90 градусов. Развитием данного направления является работа [25], в которой решается задача активного слежения за объектом с динамическим изменением топологии группы разнородных агентов, имеющих различные сенсоры. Агенты вычисляют градиент неопределенности и совершают маневры для оптимизации угла параллакса. Локальная согласованная триангуляция на базе двух агентов может быть получена из общего случая. При статическом состоянии ведущего агента

ведомый опорный агент стремится занять оптимальную точку пространства, обеспечивающую формирование треугольника «агент-объект-агент» с прямым углом в вершине для повышения точности измерения дистанции до объекта. Существенную роль в этих алгоритмах играет колесная одометрия, что ограничивает их использование в POM-MRS. То же касается видеоцентрических SMRS, опирающихся на инерциальную одометрию для повышения точности измерения стереобазиса. Например, в работе [26] для SMRS на базе пары одинаковых агентов (БЛА) предложены алгоритмы управления распределенным стереобазисом и консенсуса, минимизирующие ошибку определения дистанции до объекта окружения.

В современных POM-MRS задачи обнаружения и выбора опорного агента для измерений дистанции до объекта окружения решаются с помощью комплексное нейросетевое детектирование агентов. В работе [5] колесный робот-агент определяет дистанцию до другого робота по визуальным маркерам на его корпусе. Выбор опорного робота для измерения дистанции до объекта окружения происходит с учетом угла и расстояния до другого робота (чем меньше, тем лучше с точки зрения нейросетевой видеоаналитики). Из-за малых размеров маркеров рабочая дистанция между роботами составляет 1 – 5 м для распознавания и 8 – 10 м для обнаружения.

В [9] используется нейросетевое распознавание и предсказание относительного масштаба и ориентации агента-БЛА по ключевым точкам корпуса, после чего дистанция до опорного агента вычисляется по известному размеру топологии этих ключевых точек в соответствующем ракурсе. Распознавание требует высококачественного изображения БЛА и работает только при расстоянии между БЛА до 4 м (при использовании широкоугольного объектива), а при большем удалении дистанция вычисляется без учета ключевых точек по среднему размеру ограничивающей рамки, налагаемой на изображение БЛА нейросетевым детектором. БЛА стремятся поддерживать дистанцию, обеспечивающую минимум риска столкновения при максимуме стереобазиса в ущерб точности определения дистанции до объектов окружения. Для оценки дистанции до объектов окружения используется нейросетевое распознавание объектов и динамическая триангуляция на основе данных инерциальной системы БЛА. При этом реализуется принцип динамического выбора опорного агента с учетом углов и дистанций без непосредственной корректировки местоположения роботов.

Расстояние между колесными роботами на плоскости в [8] вычисляется на основе проективной геометрии по известной ширине робота с геометрической коррекцией ракурса по визуальным маркерам, которые используются также для нейросетевого распознавания робота. При этом накладывается жесткое ограничение на топологию системы – относительное положение роботов должно отвечать равносторонней формации. Дистанция до объекта вычисляется на основе триангуляции, опираясь на концепцию оптимальной стереобазы, для реализации которой роботы поддерживают определенную дистанцию (не более 7 м), гарантирующее надежное нейросетевое распознавание. Данные инерциальной системы робота используются в фильтре Калмана для сглаживания визуальных измерений и предсказания положения робота при пропуске кадров.

Реализация POM-MRS на основе комплексного нейросетевого детектирования агентов рассматривается также в работе [23], но вместо согласованной триангуляции в ней реализуется непосредственная оценка дистанций нейронной сетью. В POM-MRS на основе комплексного нейросетевого детектирования агентов вводится жесткое ограничение на максимальный стереобазис, который с учетом угла обзора видеокамеры не превышает 7 м. Агенты стремятся поддерживать такой стереобазис для повышения точности определения дистанции до объекта окружения. Однако, согласно выражению (1) при увеличении дистанции до объекта окружения для повышения точности ее определения стереобазис также необходимо увеличивать. При этом, согласно выражению (2) для повышения точности определения самого стереобазиса его необходимо уменьшать (ошибка определения стереобазиса возрастает квадратично от стереобазиса). Из-за данного противоречия оптимальный стереобазис, обеспечивающий минимум ошибки определения дистанции до объекта окружения, зависит от этой дистанции (рис. 3) и в условиях открытых пространств, когда объекты окружения удалены от POM-MRS на сотни метров, может быть существенно больше 7 м (около 30 м для дистанции 1000 м).

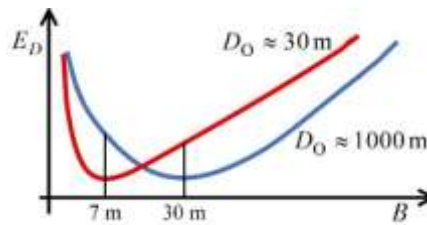


Рис. 3. Зависимость ошибки E_D определения дистанции до объекта окружения от стереобазиса B для двух значений дистанции D_O до объекта окружения

Определение дистанции до объекта окружения в известных концепциях POM-MRS опирается на следующие базовые функции: 1) сетевая идентификация агентов (установление соответствия между изображением агента, его относительными координатами и сетевым адресом); 2) определение распределенного стереобазиса (использование данных об известных размерах агентах для вычисления дистанции до опорного агента); 3) вычисление дистанции до объекта окружения (на основе триангуляции); 4) согласованная триангуляция для уменьшения ошибки определения дальности до объекта окружения (геометрическая оптимизация стереобазиса, реализуемая через выбор опорного агента и масштабирование топологии сети глобально, в группе или локально через и перемещение опорного агента). Все эти функции опираются на базовую функцию передачи визуальных данных (или результатов их обработки, т.е. признакового описания или дескрипторов, или результатов их интеллектуального анализа, т.е. семантического содержания) и метаданных (параметров сетевой синхронизации агентов: пространственной синхронизации, описывающих параметры калибровки и ориентирования камеры; временной синхронизации, т.е. моментов времени формирования визуальных данных; компенсации движения агентов, позволяющих прогнозировать и отслеживать изменение положения и ориентации агентов). На рис. 4 приведены модели определения дистанции до объекта окружения в зависимости от режима работы POM-MRS, опирающиеся на рассмотренные функции.



Рис. 4. Модели определения дистанции до объектов окружения в POM-MRS

Модель определения дистанции до объекта окружения на основе локальной согласованной триангуляции

Рассматривается децентрализованная однородная POM-MRS, состоящая из двух и более агентов, имеющих одинаковую конструкцию и размеры. Предполагается, что функционирование POM-MRS осуществляется в квазистационарном режиме, при котором фазы оптимизации стереобазиса для построения 3D-карты окружения сменяются длительными периодами локальных измерительных операций в фиксированных точках, осуществляемых параллельно с основными функциями, соответствующими целевому назначению POM-MRS (например, зондирование почвы; замеры концентраций газов или радиации; поиск микротрещин в конструкциях; формирование геодезической карты и т.д.).

Разработана модель DeepGV (Distance Error Estimation and Protocol-Geometric Verification) определения дистанции до объекта окружения в однородной децентрализованной POM-MRS на основе локальной согласованной триангуляции с аналитической оценкой ошибок определения дистанции и протоколно-геометрической верификацией гипотез о детектировании агентов. Модель описывает 4 базовые функции, позволяющие построить общий алгоритм определения дистанции: определение распределенного стереобазиса на основе метода обратной проекции; вычисление дистанции до объекта окружения на основе триангуляции; локальная согласованная триангуляция, включающая выбор опорного агента (статическая оптимизация распределенного стереобазиса) и управляемое перемещение опорного агента (динамическая оптимизация распределенного стереобазиса) на основе аналитической оценки ошибок определения дистанции; локальная визуально-протокольная идентификация агентов на основе протоколно-геометрической верификации гипотез о детектировании агентов.

На основе описанных моделью DeepGV базовых функций разработан алгоритм определения дистанции до объекта окружения агентом i (рис. 5). В алгоритме базовые функции дополнены функциями поворота агента i , перемещения агента j , проверки условий прекращения вращения агента i и перемещения агента j .

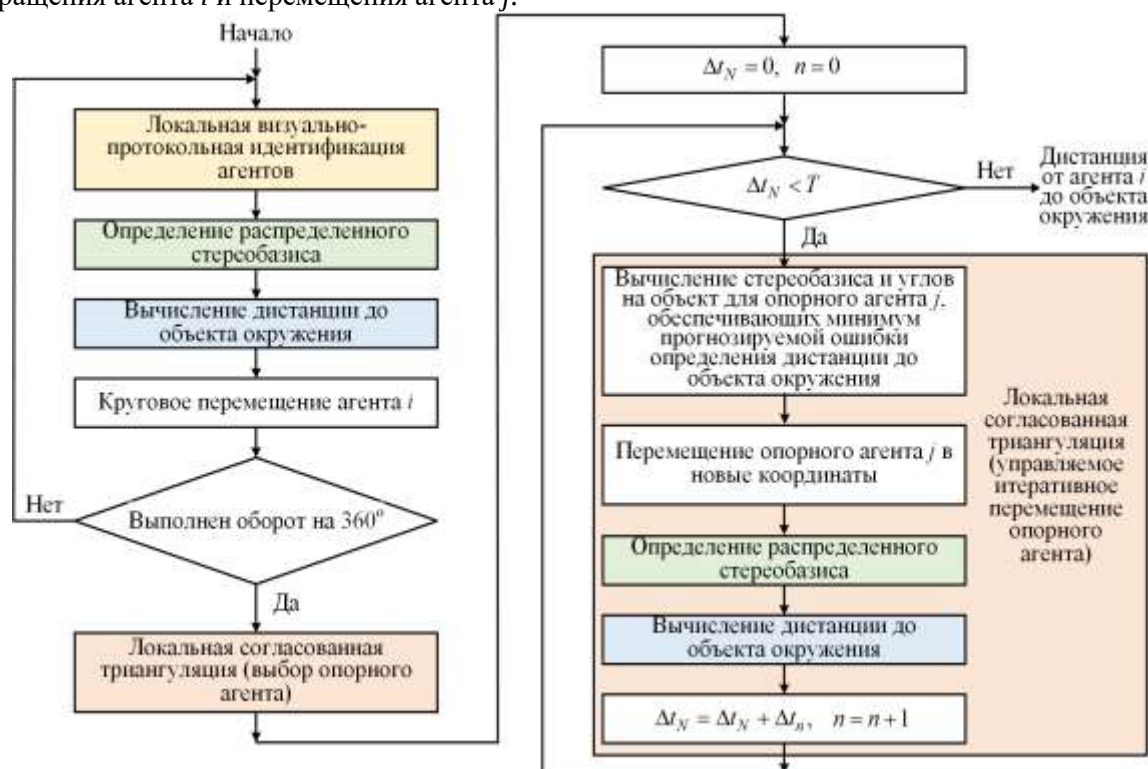


Рис. 5. Модели определения дистанции до объектов окружения в POM-MRS

В момент времени t_0 возникает необходимость установления соответствия между номерами (сетевыми адресами) агентов в окружении агента i с их относительными положениями и ориентациями, регистрируемыми по изображению $I_R(i, j, t_0)$. Для решения данной задачи

предлагается алгоритм PGHV (Protocol-Geometric Hypotheses Verification) локальной визуально-протокольной идентификации агентов, основанный на протоколно-геометрической верификации гипотез о детектировании агентов.

Оценка эффективности моделей определения дистанции до объекта окружения

Для оценки эффективности моделей определения дистанции до объекта окружения разработано программное средство, в котором стереобазис учитывается с ошибкой, зависящей от типа одометрии для SRS (см. табл., параметр T_O) или от дистанции между агентами для POM-MRS. Программное средство позволяет задать (рис. 6): для SRS – дистанцию $D_O(t_1)$ от робота до объекта, угол $\alpha_O(t_1)$ на объект в момент времени t_1 , начальное, конечное значения и шаг изменения стереобазиса $B(t_1, t_2)$ для получения оценки $D_O(t_1)$ дистанции до объекта в момент времени t_2 согласно выражению (1), значение δ относительной погрешности измерения перемещения робота с помощью одометрии различного тип; для POM-MRS – дистанцию $D(i, j, t_0)$ от агента i до объекта, угол $\alpha(i, j, t_0)$ на объект для агента i относительно направления на агент j , начальное, конечное значения и шаг изменения распределенного стереобазиса $B(i, j, t_0)$ между агентами i и j для получения оценки $D(i, j, t_0)$ дистанции до объекта от агента i в момент времени t_0 , априорную информацию о размере $S_R(A_C)$ агента вдоль измерительной оси A_C (условно лежит в горизонтальной плоскости и плоскости, параллельной плоскости фотоприемной матрицы агента i), угловой $\alpha_M(A_O(i, j, t_0))$ и пиксельный $P_M(A_O(i, j, t_0))$ размеры фотоприемной матрицы вдоль A_C .

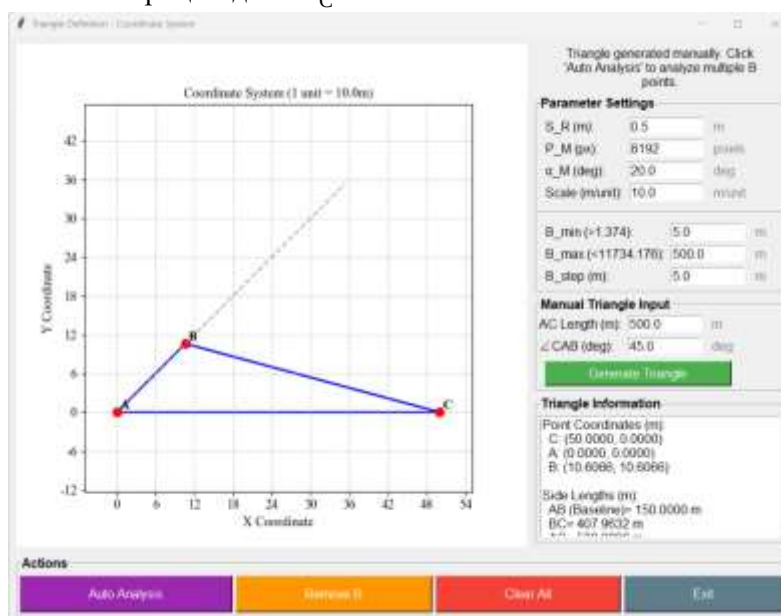


Рис. 6. Графический интерфейс программного средства оценки эффективности моделей определения дистанции до объекта окружения

В случае POM-MRS программное средство работает по следующему алгоритму. Для каждого циклически выбираемого с заданным шагом значения распределенного стереобазиса $B(i, j, t_0)$ на основе заданных значений $S_R(A_C)$, $\alpha_M(A_O(i, j, t_0))$, $P_M(A_O(i, j, t_0))$ вычисляются: с помощью выражения $\beta(i, j, t_0) = \arctan(S_R(A_C)/B(i, j, t_0))$ угловой размер $\beta(i, j, t_0)$; с помощью выражения $P_R(i, j, t_0) = \lceil \beta(i, j, t_0)/R(A_C) \rceil$ оценка $P_R(i, j, t_0)$ пиксельного

размера агента j вдоль оси A_C , где $R(A_C) = \alpha_M(A_C)/P_M(A_C)$; $[]$ – округление до ближайшего целого. Эти значения используются затем для вычисления оценки $B(i, j, t_0)$ распределенного стереобазиса с помощью выражения $B(i, j, t_0) = S_R(A_C)/\tan(\beta(i, j, t_0))$ и оценок $B_+(i, j, t_0)$ и $B_-(i, j, t_0)$ отклонений распределенного стереобазиса с помощью выражений $B_+(i, j, t_0) = S_R(A_C)/\tan(\beta_+(i, j, t_0))$ и $B_-(i, j, t_0) = S_R(A_C)/\tan(\beta_-(i, j, t_0))$, где $\beta(i, j, t_0) = P_R(i, j, t_0)R(A_C)$; $\beta_+(i, j, t_0) = (P_R(i, j, t_0) + 1)R(A_C)$; $\beta_-(i, j, t_0) = (P_R(i, j, t_0) - 1)R(A_C)$. На основе полученных значений вычисляются оценки номинальной $D(i, j, t_0)$, минимальной $D_-(i, j, t_0)$ и максимальной $D_+(i, j, t_0)$ дистанций от агента i до объекта. Затем вычисляется оценка $E_D(i, j, t_0)$ максимальной ошибки определения дистанции до объекта. Используемые в выражении (6) отклонения $\alpha_+(i, j, t_0)$, $\alpha_-(i, j, t_0)$, $\alpha_+(j, i, t_0)$, $\alpha_-(j, i, t_0)$ от номинальных значений $\alpha(i, j, t_0)$ и $\alpha(j, i, t_0)$ углов на объект вычисляются с помощью выражений $\alpha_+(i, j, t_0) = (P_A(i, j, t_0) + 1)R(A_C)$, $\alpha_-(i, j, t_0) = (P_A(i, j, t_0) - 1)R(A_C)$, $\alpha_+(j, i, t_0) = (P_A(j, i, t_0) + 1)R(A_C)$, $\alpha_-(j, i, t_0) = (P_A(j, i, t_0) - 1)R(A_C)$, где $P_A(i, j, t_0) = [\alpha(i, j, t_0)/R(A_C)]$; $P_A(j, i, t_0) = [\alpha(j, i, t_0)/R(A_C)]$; $\alpha(i, j, t_0)$ – вводится в качестве исходных данных; $\alpha(j, i, t_0) = \arcsin(D(i, j, t_0)\sin(\alpha(i, j, t_0))/D(j, i, t_0))$;

$D(j, i, t_0) = \sqrt{B(i, j, t_0)^2 + D(i, j, t_0)^2 - 2B(i, j, t_0)D(i, j, t_0)\cos(\alpha(i, j, t_0))}$. Затем на основе заданного значения $D(i, j, t_0)$ вычисляется максимальная ошибка $E_D(i, j, t_0)$ определения дистанции от агента i до объекта с помощью выражения $E_D(i, j, t_0) = \max(|D(i, j, t_0) - D_-(i, j, t_0)|, |D_+(i, j, t_0) - D(i, j, t_0)|)$.

В случае SRS программное средство работает по следующему алгоритму. Для каждого циклически выбираемого с заданным шагом значения стереобазиса $B(t_1, t_2)$ на основе значения δ относительной погрешности измерения перемещения робота с помощью одометрии различного типа вычисляются отклонения $B_+(t_1, t_2)$ и $B_-(t_1, t_2)$ стереобазиса с помощью выражений $B_+(t_1, t_2) = (1 + \delta)B(t_1, t_2)$ и $B_-(t_1, t_2) = (1 - \delta)B(t_1, t_2)$. Затем на основе этих значений вычисляются оценки номинальной $D_O(t_1)$, минимальной $D_{O-}(t_1)$ и максимальной $D_{O+}(t_1)$ дистанций от робота до объекта. Затем вычисляется оценка $E_D(t_1)$ максимальной ошибки определения дистанции с помощью выражения $E_D(t_1) = \max(|D_{O+}(t_1) - D_O(t_1)|, |D_O(t_1) - D_{O-}(t_1)|)$. Значения $D_O(t_1)$, $D_{O+}(t_1)$, $D_{O-}(t_1)$ вычисляются аналогично $D(i, j, t_0)$, $D_-(i, j, t_0)$, $D_+(i, j, t_0)$ для соответствующим значениям для POM-MRS. Затем на основе заданного значения $D_O(t_1)$ вычисляется максимальная ошибка $E_D(t_1)$ определения дистанции от робота до объекта с помощью выражения $E_D(t_1) = \max(|D_O(t_1) - D_{O-}(t_1)|, |D_{O+}(t_1) - D_O(t_1)|)$. Для инерциальной и механической

одометрии $E_D(t_1) \approx E_D(t_1)$. В случае колесной одометрии из-за прокручивания колес отклонение $B_-(t_1, t_2)$ не учитывается.

Для POM-MRS, использующей предложенную модель DeepGV определения дистанции до объекта окружения на основе локальной согласованной триангуляции с аналитической оценкой ошибок измерений и протоколно-геометрической верификацией гипотез о детектировании агентов, и однороботных систем WM, WV, WIm, WIme, QV (обозначения сформированы по параметрам $T_R | T_O$, см. табл.), на рис. 7,а приведены зависимости оценок $E_D(i, j, t_0)$, $E_D(t_1)$ ошибок (сплошные линии) и ошибки $E_D(i, j, t_0)$ (синяя пунктирная линия) от стереобазиса при $S_R(A_C) = 0,5 \text{ м}$; $\alpha_M(A_C) = \alpha_M(A_C(t_0)) = \alpha_M(A_O(i, j, t_0)) = 20^\circ$; $P_M(A_C) = P_M(A_C(t_0)) = P_M(A_O(i, j, t_0)) = 8192$ пикселей; $D_O(t_1) = D(i, j, t_0) = 500 \text{ м}$; $\alpha_O(t_1) = \alpha(i, j, t_0) = 45^\circ$ (разрешение высокое, угол на объект 45° , дистанция 500 м). Из рис. 7,а следует, что для POM-MRS на основе DeepGV существует оптимальный распределенный стереобазис, обеспечивающий минимум оценки $E_D(i, j, t_0)$ ошибки и самой ошибки $E_D(i, j, t_0)$. Для указанных условий минимальная ошибка определения дистанции до объекта обеспечивается при распределенном стереобазисе между агентами $B(i, j, t_0) \approx 25 \text{ м}$.

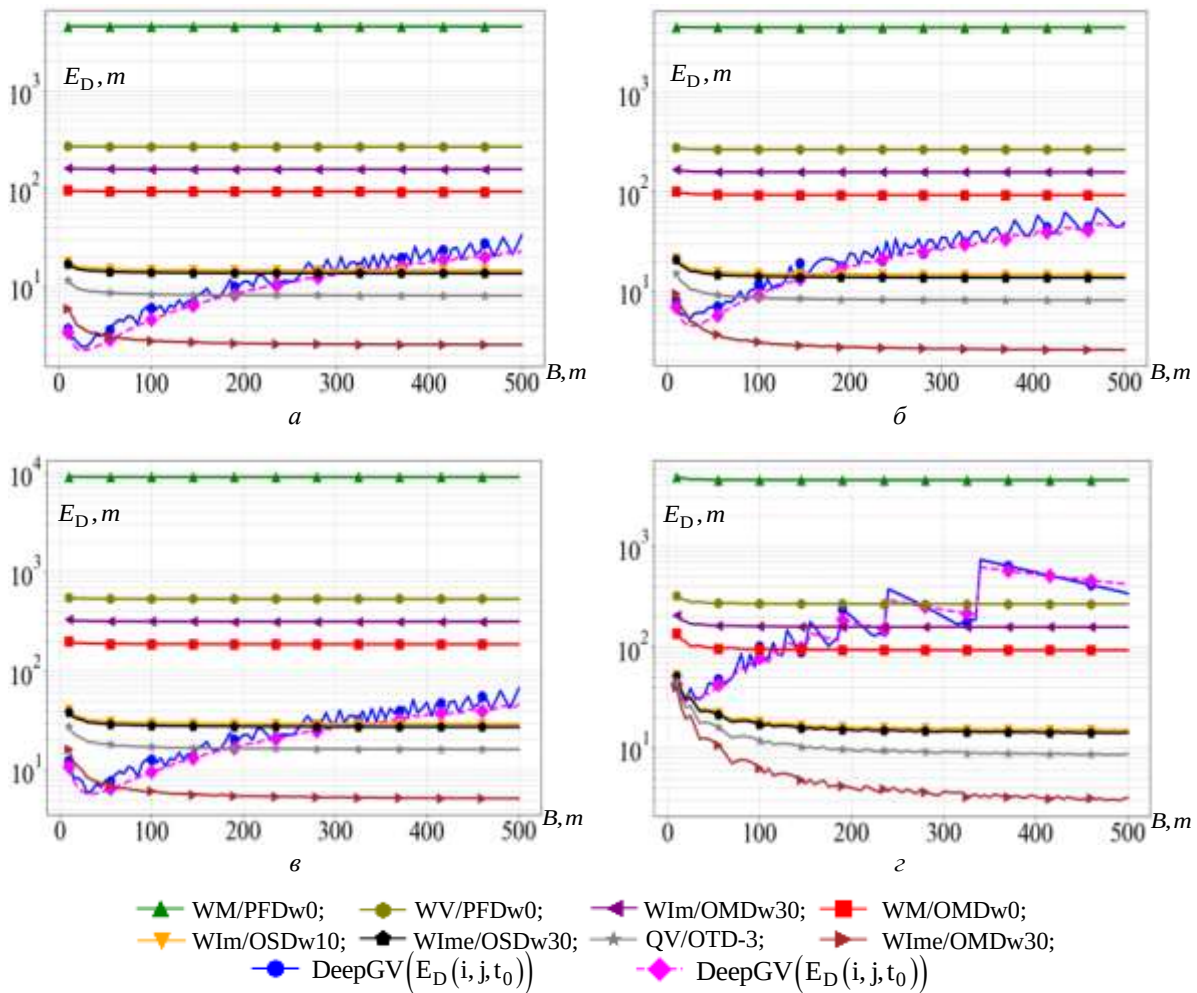


Рис. 7. Зависимости оценки ошибки определения дистанции до объекта окружения от стереобазиса: а – разрешение высокое, угол на объект 45° , дистанция 500 м; б – разрешение низкое, угол на объект 45° , дистанция 500 м; в – разрешение высокое, угол на объект 60° , дистанция 500 м; г – разрешение низкое, угол на объект 60° , дистанция 1000 м

В данной точке POM-MRS обеспечивает наименьшую по сравнению с SRS из табл. оценку ошибки $E_D(i, j, t_0) \approx 2,2 \text{ м}$ ($E_D(i, j, t_0) \approx 2,4 \text{ м}$), которая примерно в 1,7, 4,2, 6,7, 7,1, 41,5, 71,5, 120,6, 2034,4 раз меньше по сравнению с оценками ошибок для систем WIme/OMDw30, QV/OTD-3, WIme/OSDw30, WIm/OSDw10, WM/OMDw0, WIm/OMDw30, WV/PFDw0, WM/PFDw0 (см. табл.). По сравнению с POM-MRS на основе комплексного нейросетевого детектирования агентов [6, 8–11] при максимальном стереобазисе 7 м ($E_D(i, j, t_0) \approx 4,6 \text{ м}$, $E_D(i, j, t_0) \approx 5,1 \text{ м}$) POM-MRS на основе модели DeepGV позволяет уменьшить оценку $E_D(i, j, t_0)$ ошибки примерно в 2,1 раза.

При уменьшении разрешения видекамеры в 2 раза (см. рис. 7,б) минимальная оценка ошибки определения дистанции до объекта в POM-MRS на основе DeepGV увеличивается также в 2 раза и становится соизмеримой с оценкой ошибки в системе WIme/OMDw30 (разрешение низкое, угол на объект 45° , дистанция 500 м). По сравнению с POM-MRS на основе комплексного нейросетевого детектирования агентов [6, 8–11] при максимальном стереобазисе 7 м ($E_D(i, j, t_0) \approx 9,3 \text{ м}$, $E_D(i, j, t_0) \approx 10,5 \text{ м}$) POM-MRS на основе модели DeepGV позволяет уменьшить оценку $E_D(i, j, t_0)$ ошибки примерно в 2,1 раза.

При $S_R(A_C) = 0,5 \text{ м}$; $\alpha_M(A_C) = \alpha_M(A_C(t_0)) = \alpha_M(A_O(i, j, t_0)) = 20^\circ$;
 $P_M(A_C) = P_M(A_C(t_0)) = P_M(A_O(i, j, t_0)) = 8192$ пикселей; $D_O(t_1) = D(i, j, t_0) = 500 \text{ м}$;
 $\alpha_O(t_1) = \alpha(i, j, t_0) = 60^\circ$ (разрешение высокое, угол на объект 60° , дистанция 500 м) POM-MRS на основе DeepGV обеспечивает наименьшую по сравнению с системами из табл. оценку ошибки $E_D(i, j, t_0) \approx 2,0 \text{ м}$ ($E_D(i, j, t_0) \approx 2,2 \text{ м}$) при $B(i, j, t_0) \approx 25 \text{ м}$, которая примерно в 1,8, 4,5, 7,2, 7,7, 45,4, 78,3, 132,1, 2230,5 раз меньше по сравнению с оценками ошибок для систем WIme/OMDw30, QV/OTD-3, WIme/OSDw30, WIm/OSDw10, WM/OMDw0, WIm/OMDw30, WV/PFDw0, WM/PFDw0 рассмотренных в табл. 1 (см. рис. 7,с). По сравнению с POM-MRS на основе комплексного нейросетевого детектирования агентов [6, 8–11] при максимальном стереобазисе 7 м ($E_D(i, j, t_0) \approx 3,8 \text{ м}$, $E_D(i, j, t_0) \approx 4,4 \text{ м}$) POM-MRS на основе DeepGV позволяет уменьшить оценку $E_D(i, j, t_0)$ ошибки примерно в 1,9 раза.

При использовании широкоугольного объектива $\alpha_M(A_C) = \alpha_M(A_C(t_0)) = \alpha_M(A_O(i, j, t_0)) = 70^\circ$ и матрицы с меньшим пиксельным размером $P_M(A_C) = P_M(A_C(t_0)) = P_M(A_O(i, j, t_0)) = 2048$ (т.е. снижение разрешения в 14 раз) для $D_O(t_1) = D(i, j, t_0) = 1000 \text{ м}$ (при увеличении дистанции до объекта в 2 раза); $\alpha_O(t_1) = \alpha(i, j, t_0) = 60^\circ$ (разрешение низкое, угол на объект 60° , дистанция 1000 м) предложенная модель обеспечивает минимальное значение ошибки $E_D(i, j, t_0) \approx 83,5 \text{ м}$ ($E_D(i, j, t_0) \approx 87,5 \text{ м}$) при $B(i, j, t_0) \approx 35 \text{ м}$, что в 1,9, 1,6, 1,3, 1,3 раз хуже по сравнению с оценками ошибок для моделей одометрии из WIme/OMDw30, QV/OTD-3, WIme/OSDw30, WIm/OSDw10, но в 2,7, 4,4, 7,1, 111,3 раз лучше по сравнению с оценками ошибок для моделей одометрии из WM/OMDw0, WIm/OMDw30, WV/PFDw0, WM/PFDw0 (см. рис. 7,д). По сравнению с POM-MRS на основе комплексного нейросетевого детектирования агентов [6, 8–11] при максимальном стереобазисе 7 м ($E_D(i, j, t_0) \approx 248,3 \text{ м}$, $E_D(i, j, t_0) \approx 235,8 \text{ м}$) POM-MRS на основе DeepGV позволяет уменьшить оценку $E_D(i, j, t_0)$ ошибки примерно в 3,0 раза.

Заключение

Для повышения точности определения дистанции до объекта в POM-MRS, функционирующей в квазистационарном режиме, разработана модель на основе локальной согласованной триангуляции с аналитической оценкой погрешности измерений и протоколно-геометрической верификацией гипотез о детектировании агентов. Модель отличается от моделей SRS на основе инерциальной, механической, визуальной и комбинированной одометрии, а также от известных моделей POM-MRS на основе комплексного нейросетевого детектирования агентов использованием: протоколно-геометрической верификации гипотез о детектировании агентов для их локальной визуально-протокольной идентификации; аналитической оценки ошибок определения дистанции до объекта окружения при выборе опорного агента для статической оптимизации распределенного стереобазиса и при перемещении опорного агента для динамической оптимизации распределенного стереобазиса. На основе предложенной модели разработано программное средство, в котором стереобазис учитывается с ошибкой, зависящей от типа одометрии для SRS или от дистанции между агентами для POM-MRS. Показано, что предложенная модель позволяет уменьшить ошибки определения дистанции до объекта окружения по сравнению с однороботными моделями на основе одометрии различного вида и известными моделями POM-MRS с комплексным нейросетевым детектированием агентов.

Благодарность. Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (БРФФИ) и Национального фонда естественных наук Китая (НФЕНК) в рамках проекта T25КИ-054 «Технология многокамерной локализации и построения карты открытого пространства».

MODEL FOR DETERMINING THE DISTANCE TO AN OBJECT IN A QUASI-STATIONARY MULTI-ROBOT SYSTEM WITHOUT ODOMETRY BASED ON A CONSISTENT TRIANGULATION

Q. GUO, M. I. ZORKO, S.V. LITVINOV, V.Yu. TSVIATKOU

Abstract. The article considers models for determining the distance to an environmental object by a single robot and in a decentralized self-organizing video-centric multi-robot system without odometry (POM-MRS) – operating in a quasi-stationary mode. To improve the accuracy of determining the distance to an environmental object in POM-MRS, a model has been developed based on local consistent triangulation with an analytical assessment of measurement errors and protocol-geometric verification of hypotheses about agent detection. The model describes the basic functions that make it possible to construct a general algorithm for determining the distance: determination of a distributed stereo basis; calculation of the distance to an environmental object; local consistent triangulation, including the selection of a reference agent (static optimization of the local distributed stereo basis) and controlled movement of the reference agent (dynamic optimization of the distributed stereo basis); local visual-protocol identification of agents based on protocol-geometric verification of hypotheses about their detection. It is shown that the proposed model allows to reduce errors in determining the distance to the surrounding object in comparison with single-robot models based on various types of odometry and well-known POM-MRS models with complex neural network detection of agents.

Keywords: object distance determination, video-centric self-organizing multi-robot system without odometry.

Список литературы

1. Mur-Artal R., Montiel J. M. M., Tardos J. D. (2015) ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*. 31(5), 1147–1163.
2. Mur-Artal R., Tardós J. D. (2017) ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*. 33(5), 1255–1262.
3. Rodríguez-Martínez E. A., Flores-Fuentes W., Achakir F., Sergiyenko O., Murrieta-Rico F. N. (2025) Vision-Based Navigation and Perception for Autonomous Robots: Sensors, SLAM, Control Strategies, and Cross-Domain Applications—A Review. *Eng.* 6(7), 153.

4. Chen W., Wang X., Gao S., Shang G., Zhou C., Li Z., Xu C., Hu K. (2023) Overview of Multi-Robot Collaborative SLAM from the Perspective of Data Fusion. *Machines*. 11(6), 653.
5. Jang Y., Oh C., Lee Y., Kim H. J. (2021) Multirobot collaborative monocular SLAM utilizing rendezvous. *IEEE Transactions on Robotics*. 37(5), 1469–1486.
6. Schmuck P., Chli M. (2017) Multi-UAV collaborative monocular SLAM. *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3863–3870.
7. Wang S., Wang Y., Li D., Zhao Q. (2023) Distributed Relative Localization Algorithms for Multi-Robot Networks: A Survey. *Sensors*. 23(5), 2399.
8. Kang T., Kwon J.-W., Bae I., Kim J. H. (2025) Monocular Vision-Based Swarm Robot Localization Using Equilateral Triangular Formations. *Machines*. 13(8), 667.
9. Si X., Xu G., Ke M., Zhang H., Tong K., Qi F. (2023) Relative Localization within a Quadcopter Unmanned Aerial Vehicle Swarm Based on Airborne Monocular Vision. *Drones*. 7(10), 612.
10. Mezey D., Bastien R., Zheng Y., et al. (2025) Purely vision-based collective movement of robots. *npj Robot.* 3(1), 11.
11. Krížek M., Vrba M., Kulaš A. B., Bogdan S., Saska M. (2024) Bio-inspired visual relative localization for large swarms of UAVs. *2024 IEEE International Conference on Robotics and Automation (ICRA)*. Yokohama, Japan, 11825–11831.
12. Chen S., Mei S., Jia G., Li Y., Zhao W. (2021) KFPA monocular ranging algorithm design and application in mobile edge computing. *Journal of Internet Technology*. 22(5), 1131–1142.
13. Alsadik B. (2016) A modified method for image triangulation using inclined angles. *23rd International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Congress, ISPRS 2016: From human history to the future with spatial information. International Society for Photogrammetry and Remote Sensing (ISPRS)*. 453–458.
14. Yu Y., Pradalier C., Zong G. (2011) Appearance-based monocular visual odometry for ground vehicles. *2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. 862–867.
15. Piyathilaka L., Munasinghe R. (2010) An experimental study on using visual odometry for short-run self localization of field robot. *2010 Fifth International Conference on Information and Automation for Sustainability*. IEEE. 150–155.
16. Basiuk V., Maksymova S., Chala O., Miliutina O. (2023) Mobile robot position determining using odometry method. *Multidisciplinary Journal of Science and Technology*. 3(3), 227–234.
17. Paryanto P., Pratama R. R., Saputra R. P. (2023) Wheel odometry-based localization for autonomous wheelchair. *2023 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*. IEEE. 357–362.
18. Ojeda L., Cruz D., Reina G., Borenstein J. (2006) Current-based slippage detection and odometry correction for mobile robots and planetary rovers. *IEEE Transactions on Robotics*. 22(2), 366–378.
19. Strydom R., Thurrowgood S., Srinivasan M. V. (2014) Visual odometry: autonomous UAV navigation using optic flow and stereo. *Australasian Conference on Robotics and Automation, ACRA. Australian Robotics and Automation Association*. 2.
20. Reddy B. L. (2019) Evolution & development of visual odometry systems. *Anveshana's International Journal of Research in Engineering and Applied Sciences*. 4(12). 64–76.
21. Shi C. X., Yang G. H. (2023) Secure bearing-based target localization for multi-agent networks against malicious agents. *IEEE Transactions on Automation Science and Engineering*. 21(4), 5812–5825.
22. Ding Y., Qin X., Cong Y., Wang X. (2026) Set-Membership Estimation Based Distributed Cooperative Localization of Multiple UAVs. *IEEE Transactions on Signal and Information Processing over Networks*. 12, 248–259.
23. Rocchi A., Wang Z., Pan Y.-J. (2023) A Practical Vision-Aided Multi-Robot Autonomous Navigation using Convolutional Neural Network. *2023 IEEE 6th International Conference on Industrial Cyber-Physical Systems (ICPS)*. Wuhan, China, 1–6.
24. Bishop A. N., et al. (2010) Optimality Analysis of Sensor-Target Localization Geometries. *IEEE Transactions on Aerospace and Electronic Systems*. 46(2), 779–796.
25. Zhou K. X., Roumeliotis S. I. (2011) Multirobot Active Target Tracking with Combinations of Relative Observations. *IEEE Transactions on Robotics*. 27(4), 678–695.
26. Karrer M., Chli M. (2021) Distributed variable-baseline stereo SLAM from two UAVs. *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 82–88.

ЭКСПЕРИМЕНТАЛЬНАЯ СИСТЕМА ДЛЯ ИССЛЕДОВАНИЯ КАНАЛА СВЯЗИ ВИДИМЫМ СВЕТОМ НА БАЗЕ SDR

А.Н. СОЛОВЬЁВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 3 апреля 2026

Аннотация. В работе представлена экспериментальная система для исследования канала связи видимым светом, реализованная на базе программно-определяемых радиосистем и оптоэлектронного тракта. Предложенная установка обеспечивает формирование, передачу и прием сигналов с возможностью управляемого изменения параметров оптического канала и условий распространения. Система ориентирована на регистрацию параметров сигнала, что позволяет исследовать динамические свойства канала, что может быть использовано для исследования временной структуры канала и ее анализа.

Ключевые слова: система связи видимым светом; программно-определяемая радиосистема; OFDM; анализ сигналов; отношение сигнал/шум; временные характеристики; динамика канала.

Введение

Системы связи видимым светом (СВС) являются перспективным направлением развития беспроводных технологий передачи данных. Для организации канала связи используется видимый диапазон света от 380 до 780 нм, что позволяет совмещать применение световых источников для освещения и передачи данных [1]. Увеличение нагрузки на радиочастотный спектр делает системы СВС все более привлекательным дополнением, а в некоторых случаях альтернативой традиционным системам связи [2].

В настоящее время исследования в области систем СВС направлены на повышение надежности передачи данных и на повышение пропускной способности. Основными широко используемыми показателями оценки качества соединения являются такие показатели, как вероятность битовой ошибки (BER), отношение сигнал/шум (SNR) и связанный с ним Q фактор – показатель качества цифрового сигнала [3]. Эти параметры являются базовыми для анализа систем передачи данных и используются для выбора модуляции и алгоритмов обработки сигналов [4].

Несмотря на повсеместное применение данных показателей, их использование в основном ограничивается оценкой канала в определенный фиксированный момент времени или применяется усреднение показателя за выбранный интервал. В реальных условиях канал СВС подвергается воздействию множества факторов, таких как внешняя освещенность, отражения, геометрия распространения сигнала в помещении и специфика оптических и электронных компонентов системы. Все это приводит к формированию динамических процессов и изменению характеристик канала во времени [5].

Для изучения этих явлений требуется исследовательская экспериментальная система, которая может обеспечить формирование сигналов для передачи в системах СВС, управляемое изменение параметров канала и регистрацию динамических характеристик при наблюдении за работой системы. Использование программно определяемой радиосистемы (ПОР) в сочетании с настраиваемым оптоэлектронным трактом прохождения сигнала позволяют построить подобную систему.

В данной работе представлена система, построенная на базе ПОР в сочетании с оптоэлектронным трактом, обеспечивающая передачу и прием оптических сигналов с различными типами модуляции в диапазоне частот 1–10 МГц. Особенностью системы является возможность регистрации параметров сигнала во времени при различных внешних воздействиях, что создает основу для последующего анализа динамически характеристик канала.

Архитектура экспериментальной системы

Архитектура экспериментальной системы исследования канала СВС представляет собой тракт передачи и приема сигнала, включающий формирование сигнала, его передачу через оптический канал и последующую регистрацию и обработку на приемной стороне.

Важной особенностью предложенной системы является ее ориентация на непрерывное формирование и регистрацию сигнала и его характеристик во времени. На рис. 1 представлена архитектура экспериментальной установки. Сигнал с настраиваемыми характеристиками передается из формирователя сигнала в программно определяемую радиосистему, после чего поступает в тракт оптоэлектронного передатчика, где преобразуется из электрического в оптический. Далее сигнал распространяется по оптическому каналу и принимается оптоэлектронным приемником. Преобразованный из оптического в электрический сигнал предварительно обрабатывается приемной программно определяемой радиосистемой и передается для обработки, сохранения и анализа в обработчик сигнала.

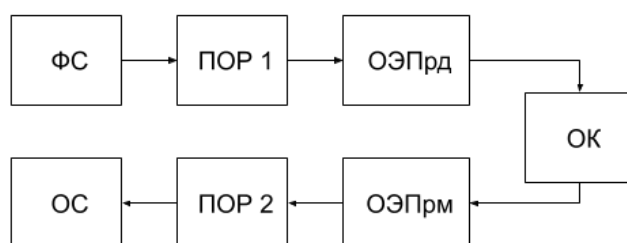


Рис. 1. Архитектура экспериментальной системы исследования канала: ФС – формирователь сигнала; ПОР 1,2 – программно определяемые радиосистемы; ОЭПрд – оптоэлектронный передатчик; ОК – оптический канал; ОЭПрм – оптоэлектронный приемник; ОС – обработчик сигнала

Экспериментальная система может быть реализована в двух основных конфигурациях, отличающихся способом организации формирования и обработки сигналов.

В первой конфигурации используются два вычислительных устройства (формирователя и обработчика сигнала), обеспечивающих раздельную работу передающего и приемного трактов. Такой подход позволяет снизить нагрузку на каждое устройство и повысить стабильность функционирования системы при длительных измерениях.

Во второй конфигурации формирование и прием сигнала могут осуществляются на одном вычислительном устройстве. Данный вариант упрощает структуру, однако предъявляет повышенные требования к вычислительным ресурсам, а также может приводить к дополнительным задержкам обработки.

Независимо от выбранной конфигурации, система ориентирована на обеспечение непрерывного формирования и обработки сигнала. Это позволяет регистрировать характеристики канала и использовать их для анализа его динамических свойств, включая изменение состояния канала под воздействием изменяемых внешних факторов.

Формирование и обработка сигнала

Формирование и прием сигналов в экспериментальной системе производятся посредством ПОР в среде GNU Radio [6], которая представляет собой открытую платформу для разработки и исследования цифровых систем связи. В среде возможно реализовать тракт приема и передачи

электрического сигнала в виде графа, состоящего из различных функциональных блоков, что обеспечивает гибкую настройку параметров сигнала и его анализ в процессе работы системы

В качестве тестовых сигналов для исследования временных характеристик канала могут использоваться как простые, так и сложные виды сигналов. Применение гармонического сигнала позволяет базово оценивать характеристики канала получая отношение сигнал/шум или изучая частотные искажения. Но для оценки канала в полной мере целесообразнее применять широкополосные сигналы используемые для передачи информации.

В качестве такого сигнала предлагается использовать сложный сигнал с ортогональным частотным разделением каналов (Orthogonal Frequency-Division Multiplexing OFDM), который широко применяется в современных системах СВЧ [7, 8]. Выбор OFDM также может быть обусловлен тем, что помимо широкого спектра он содержит поднесущие, которые могут по-разному реагировать на изменения в канале связи, что может повысить чувствительность экспериментальной установки и позволит лучше анализировать состояние канала СВЧ.

Формирование OFDM сигнала реализовано на основе стандартных блоков GNU Radio, которые были настроены под параметры экспериментальной установки. Граф передатчика, который определяет последовательность преобразований сигнала представлен на рис. 2.

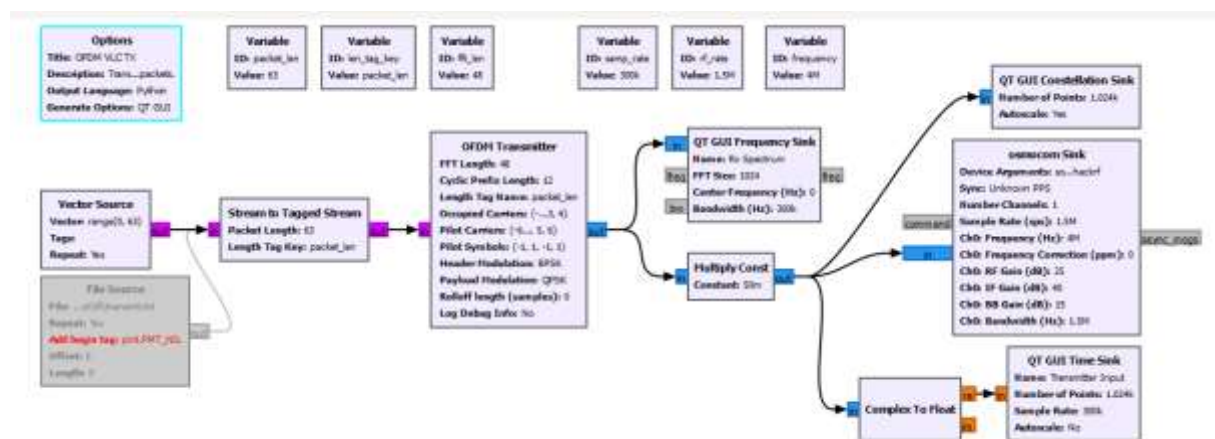


Рис. 2. Граф ФС экспериментальной установки системы СВЧ

Граф ФС представляет собой передатчик и включает в себя на выбор: один из источников данных в виде файла, либо в виде вектора (пилообразный сигнал), из которых формируется последовательность из символов либо чисел, которая далее преобразуется в поток с метками пакетов определенной длины. Для этого используется блок преобразования потока в пакетную структуру с метками (Stream to Tagged Stream), который обеспечивает корректную сегментацию данных для OFDM-модуляции.

Важнейшим элементом передающего тракта является блок OFDM-модулятора, в котором выполняется формирование поднесущих, добавление пилотных сигналов, а также преобразование сигнала с использованием быстрого преобразования Фурье (Fast Fourier Transform, FFT). С помощью переменных задаются параметры FFT, количество поднесущих и пилотных символов, а также тип модуляции полезной нагрузки и заголовка.

После формирования сигнала выполняется масштабирование его амплитуды для согласования уровней с последующим трактом передачи. Дополнительно в графе предусмотрены средства визуального контроля сигнала, включая отображение спектра осциллограммы выходного сигнала и созвездия, что позволяет контролировать корректность формирования сигнала

Результат запуска графа ФС на исполнение, можно видеть на рис. 3.

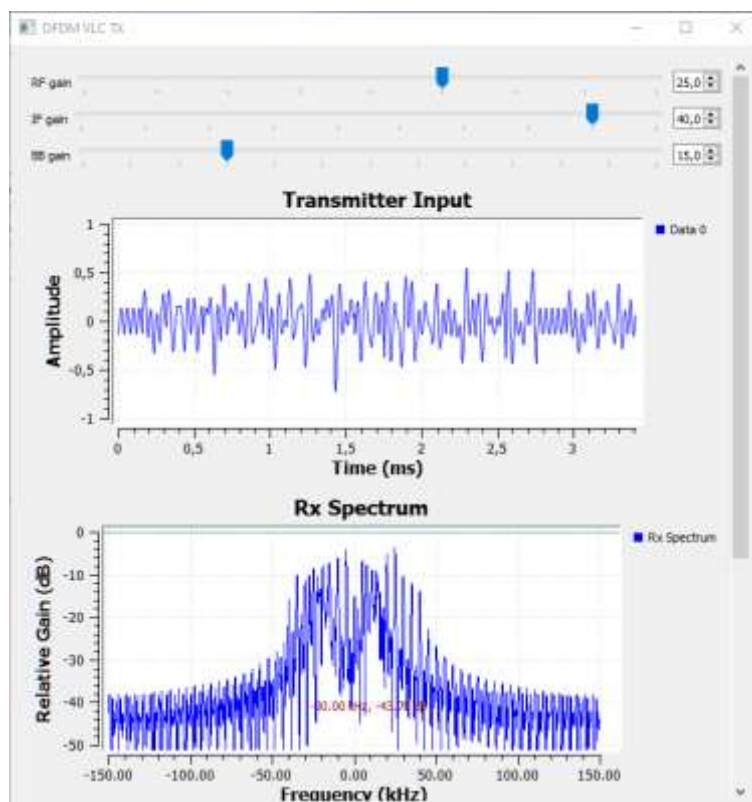


Рис. 3. Выполняемый граф ФС с параметрами усиления и выходным спектром на передачу
Граф ОС, представленный на рис. 4 реализует прием, обработку и визуализацию сигнала.

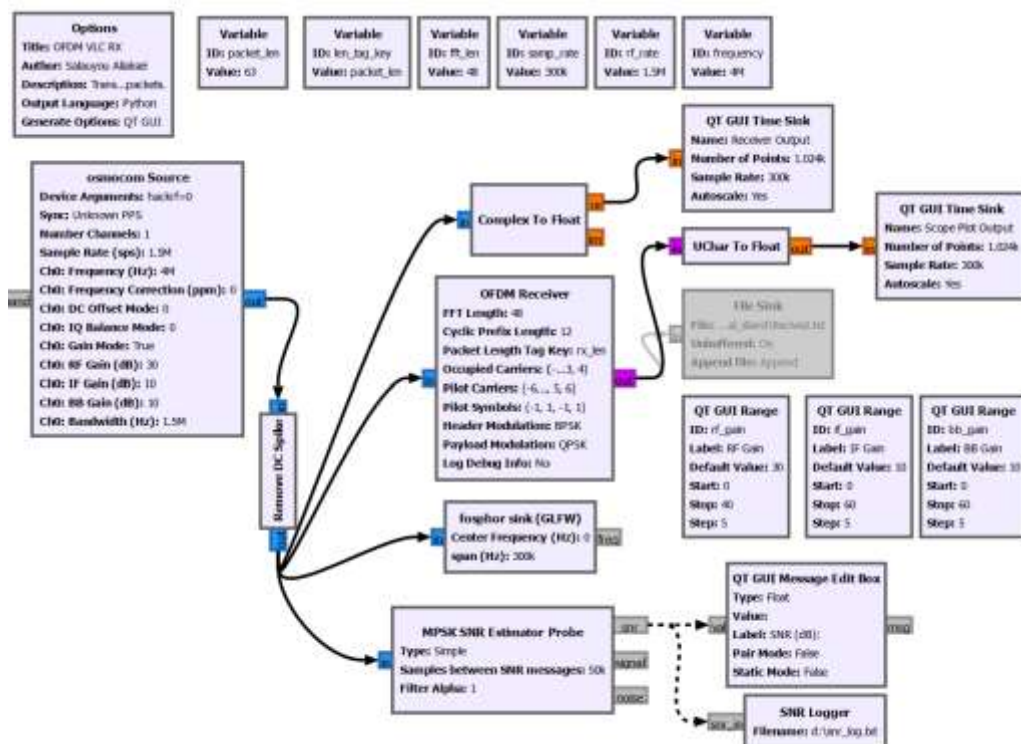


Рис. 4. Граф ОС экспериментальной установки системы CBC

Из сигнала, поступающего с приемного устройства, предварительно удаляется постоянная составляющая, которая негативно влияет на оценку канала. Далее сигнал поступает в блок OFDM-приемника, в котором выполняется синхронизация по пилот сигналам, перенос сигнала в частотную область и демодуляция поднесущих.

После демодуляции выполняется преобразование полученных данных в формат, пригодный для дальнейшей обработки и анализа, а так как основным источником данных в рамках будущей работы для нас является отношение сигнал/шум, остальные блоки вывода временно деактивированы. При необходимости отслеживания коэффициента битовой ошибки (Bit Error Rate – BER) существует возможность записи принятой последовательности в файл для дальнейшей обработки. Также в графе предусмотрены блоки визуализации временной формы сигнала, входных и выходных данных, что позволяет контролировать процесс приема и сопоставлять переданный и принятый сигналы, что особенно удобно в случае передачи векторного сигнала. Также полезен в визуализации и настройке блок «Водопад» (Waterfall), который позволяет не только отслеживать изменения сигналов в реальном времени, но и видеть историю эфира. Результат выполнения графа ОС представлен на рис. 5.

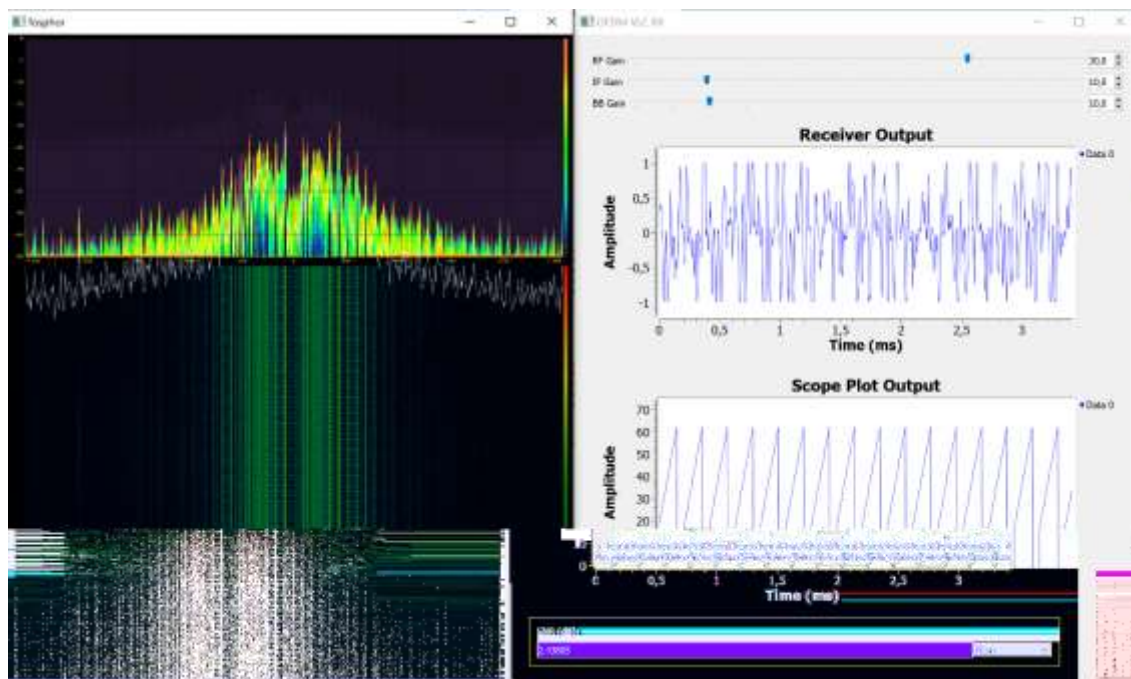


Рис. 5. Выполняемый граф ОС с параметрами усиления принимаемым спектром с памятью и принятым демодулированным сигналом

Важным элементом является блок оценки отношения сигнал/шум, используемый в приемном тракте. Встроенный модуль оценки SNR позволяет получать текущие значения качества сигнала непосредственно в процессе работы системы. Полученные значения через интерфейс пользователя записываются в файл с добавлением временной метки, что обеспечивает формирование временных рядов параметра SNR очень удобных для последующего анализа.

Реализация передающего и приемного трактов в виде графов GNU Radio обеспечивает не только передачу и прием сигнала, но и задействует встроенные средства его анализа. Это позволяет использовать систему как инструмент для исследования характеристик канала, включая их изменение во времени при воздействии различных помех.

Оптоэлектронный тракт

Оптоэлектронный тракт (ОТ) экспериментальной системы служит для преобразования сигнала из электрической формы в оптическую и обратно. В соответствии со структурной схемой на рис. 1 в ОТ можно выделить следующие основные элементы: ПОР 1; ОЭПрд; ОК; ОЭПрм; ПОР 2.

ПОР 1 и ПОР 2 представляют собой программно-определяемые радиосистемы, выполняющие функции цифро-аналогового и аналого-цифрового преобразования, а также первичной обработки сигналов. В базовой конфигурации такие системы обеспечивают формирование и прием сигналов с заданными параметрами и их передачу между цифровой и аналоговой частями тракта.

В общем случае программно-определяемые радиосистемы реализуют настраиваемую архитектуру, в которой функции обработки сигнала распределяются между аппаратной и программной частями. В некоторых устройствах для ускорения обработки используются программируемые пользователем вентильные матрицы (ППВМ, FPGA) и тогда значительная часть алгоритмов преобразования сигнала выполняется на стороне ПОР, а не вычислительной системы.

В рассматриваемой реализации в качестве ПОР используются два устройства HackRF One, которые обеспечивают формирование и прием сигналов в широком диапазоне частот [9]. Применение подобных устройств позволяет оперативно изменять параметры сигнала и проводить эксперименты без модификации аппаратной части. В качестве альтернативных решений могут рассматриваться более производительные платформы, такие как LimeSDR [10], а также другие аналогичные устройства, отличающиеся расширенной полосой пропускания и улучшенными характеристиками аналого-цифровых преобразователей. Экспериментальная установка рассчитана на такие изменения.

Использование ПОР в задачах оптической беспроводной связи давно вызывает интерес у научного сообщества и рассматривается, например, в работах [11, 12], где показана возможность построения и анализа гибких систем передачи данных на базе программно-определяемых радиосистем.

К основным преимуществам применения ПОР в составе оптоэлектронного тракта относятся:

- гибкость настройки параметров сигнала;
- возможность реализации различных методов модуляции;
- интеграция с программными средствами анализа;
- возможность регистрации параметров сигнала во времени.

К недостаткам можно отнести:

- ограниченную точность и динамический диапазон по сравнению со специализированными решениями;
- наличие дополнительных задержек, связанных с буферизацией и обработкой;
- зависимость характеристик от вычислительных ресурсов системы.

ОЭПрд обеспечивает преобразование электрического сигнала, поступающего с ПОР 1, в оптическое излучение. Управление лазерным диодом (в частном случае LPC-815 650 нм) осуществляется посредством модуляции напряжением входного сигнала источника тока. Такой способ соответствует принципу прямой модуляции интенсивности (IM/DD), широко применяемому в системах оптической связи [7]. Использование лазерного диода обеспечивает высокую направленность излучения и достаточную мощность сигнала для проведения экспериментов.

ОК представляет собой среду распространения сигнала, параметры которой могут изменяться в процессе эксперимента. В отличие от волоконно-оптических систем, в данном случае используется свободное пространство, что приводит к дополнительным потерям и влиянию внешних воздействий.

ОЭПрм реализует обратное преобразование сигнала. Фотодиод преобразует оптическое излучение в электрический ток, который затем усиливается трансимпедансным усилителем и преобразуется из токового сигнала в пропорциональное напряжение. Основные свойства кремниевых р-і-n фотодиодов с большой площадью перехода, влияющие на пропускную способность описаны в исследованиях [13]. В экспериментальной установке прием сигнала осуществляется фотодиодом типа ФД263 [14], обладающим достаточной чувствительностью в видимом диапазоне и пригодным для построения быстродействующих фотоприемных устройств.

ОЭПрд и ОЭПрм реализованы на одной печатной плате, представляющей собой модифицированный вариант ранее разработанного Li-Fi модема [15]. В рамках данной экспериментальной установки из схемы были исключены отдельные узлы обработки, что позволило получить доступ к сигналу фотоприемника без предварительной фильтрации и коррекции. В предлагаемой системе обеспечивается возможность регистрации исходного сигнала с фотодиода, содержащего все искажения, шумы и временные изменения, возникающие в канале и элементах.

Внешний вид оптоэлектронной части установки представлен на рис. 6.



Рис. 6. Оптоэлектронная часть экспериментальной установки

Следует отметить, что используемые оптоэлектронные компоненты не являются жестко фиксированными элементами системы и могут быть заменены в зависимости от условий эксперимента. Это позволяет адаптировать установку под различные диапазоны мощностей, расстояний и условий распространения сигнала.

Таким образом, ПОР, совместно с ОТ, формируют единый преобразующий контур, в котором физические искажения и шумы непосредственно отражаются на характеристиках сигнала, доступных для последующего анализа.

Оптический канал

ОК экспериментальной системы представляет собой управляемую среду распространения оптического излучения между оптоэлектронным передатчиком и приемником формируя замкнутую петлю. В отличие от волоконно-оптических линий связи, такой канал подвержен значительному влиянию внешних факторов, что делает его удобным объектом для экспериментальных исследований с возможностью вносить помехи различного характера.

Конструктивно канал включает фокусирующую линзу и систему из двух отражающих элементов (зеркал), положение которых может изменяться в очень широком диапазоне. Такая конфигурация позволяет варьировать длину оптического пути, угол падения и отражения излучения, а также формировать различные режимы распространения сигнала.

Изменение геометрии канала приводит к изменению условий распространения сигнала, включая потери мощности, расходимость пучка и влияние отраженных компонент. В ряде работ показано, что для систем связи видимым светом характерно существенное влияние многолучевого распространения и отражений от поверхностей, что приводит к искажению сигнала [5].

На рис. 7 показан один из возможных путей прохождения сигнала, траектория которого проявляется в мелкодисперсионных частицах внесенных в канал.

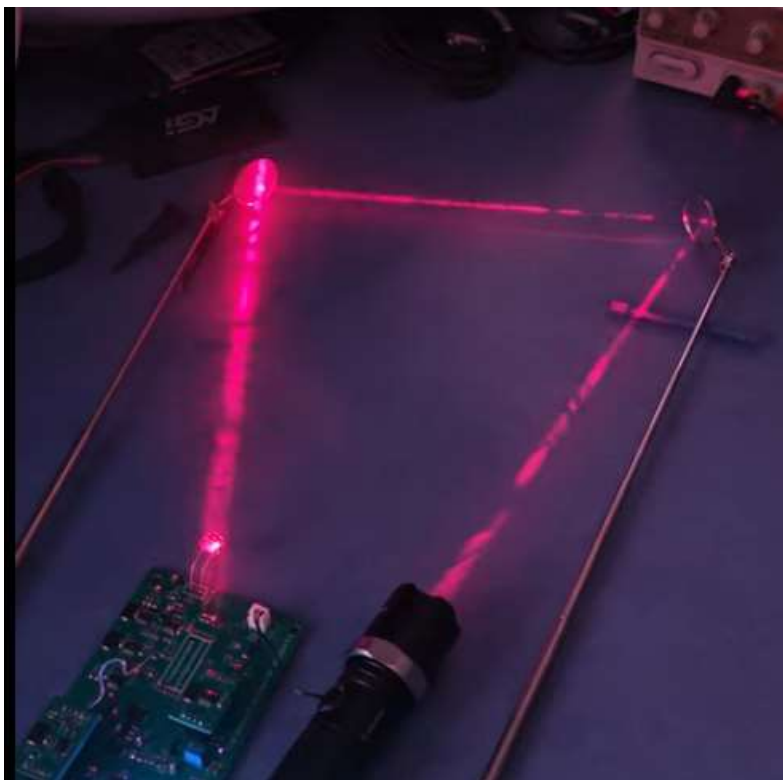


Рис. 7. Конфигурация прохождения света в мелкодисперсионной взвеси

При определенных конфигурациях оптической системы возможно формирование интерференционной картины вследствие наложения расфокусированного луча отраженного одновременно от двух зеркал. Это приводит к пространственно-временным изменениям интенсивности сигнала на приемнике, что также влияет на параметры принимаемого сигнала.

Особенностью рассматриваемой установки является возможность управляемого изменения параметров или внесения помех в канал в процессе эксперимента. Перемещение зеркал и изменение углов отражения позволяют моделировать различные сценарии распространения сигнала, включая увеличение длины канала, появление дополнительных отражений и изменение условий освещенности.

Следует отметить, что в отличие от статических моделей канала, используемых в ряде теоретических работ, реальный оптический канал в помещении является нестационарным. Его параметры изменяются во времени под воздействием внешних факторов, таких как изменение освещенности, движение объектов и нестабильность положения элементов системы.

Оптический канал в предлагаемой системе является не только средой передачи сигнала, но и активным элементом, формирующим его характеристики. Управляемая геометрия канала позволяет получать различные режимы распространения сигнала и исследовать влияние этих условий на его параметры во времени, что является ключевым для анализа его динамических свойств.

Особенности и ограничения системы

Экспериментальная система, реализующая передачу сигнала в канале СВЧ с использованием ПОР и ОТ, обладает рядом особенностей, которые оказывают влияние на характеристики принимаемого сигнала.

К числу таких особенностей относится ограниченная полоса пропускания оптоэлектронных компонентов, включая лазерный диод, фотодиод и усилительные каскады. Данные ограничения приводят к частотным искажениям сигнала и неравномерности амплитудно-частотной характеристики тракта.

Дополнительным фактором является нелинейность элементов передающего тракта, в первую очередь источника тока и лазерного диода. Нелинейные искажения приводят к изменению спектра сигнала и появлению дополнительных гармонических составляющих, что характерно для систем с прямой модуляцией излучения.

На приемной стороне существенное влияние оказывают шумы фотоприемника и усилительных каскадов, включая тепловой и дробовой шум.

Особенностью используемых ПОР является наличие задержек, связанных с буферизацией и обработкой данных, а также использование оконных методов при оценке параметров сигнала. В частности, оценка отношения сигнал/шум осуществляется на основе усреднения по выборке, что приводит к сглаживанию быстрых изменений сигнала во времени.

Следует также отметить нестационарный характер оптического канала. Изменение геометрии распространения сигнала, внешней освещенности и положения элементов установки приводит к изменению параметров канала во времени.

Перечисленные особенности не только ограничивают точность передачи данных, но и формируют наблюдаемую структуру сигнала на выходе системы. В отличие от специализированных систем связи, где подобные эффекты стремятся минимизировать, в рамках данной работы они рассматриваются как источник информации о свойствах канала.

Таким образом, ограничения экспериментальной системы одновременно являются факторами, формирующими динамику сигнала, что позволяет использовать их для исследования временных характеристик канала связи видимым светом.

Обсуждение и заключение

В работе представлена экспериментальная система для исследования канала СВС, реализованная на базе ПОР и ОТ и ОК. Установка обеспечивает формирование, передачу и прием сигнала с возможностью гибкой настройки параметров и управляемого изменения условий распространения.

В отличие от традиционных подходов, основанных на анализе усредненных характеристик, таких как SNR и BER, предложенный подход ориентирован на изучение поведения сигнала во времени. Это позволяет рассматривать канал связи как динамическую среду, параметры которой изменяются под воздействием внешних факторов и конфигурации тракта.

Применение ПОР обеспечивает гибкость формирования сигналов и их последующего анализа без изменения аппаратной части. При этом изменение параметров передачи, включая, но не ограничиваясь типом и параметрами модуляции, может рассматриваться как дополнительный фактор, влияющий на наблюдаемые характеристики сигнала во времени.

ОТ и ОК формируют множество искажений и шумов, определяющих наблюдаемые характеристики сигнала. Возможность изменения геометрии распространения, включая длину пути и условия отражения, позволяет исследовать влияние параметров среды на поведение сигнала.

В результате формируются временные зависимости характеристик сигнала, которые могут быть использованы для анализа устойчивости канала и выявления закономерностей его изменения при различных условиях передачи.

Разработанная установка может рассматриваться как экспериментальная платформа для исследования динамических свойств канала связи видимым светом. Полученные результаты создают основу для дальнейшего перехода к количественным методам анализа временных зависимостей характеристик сигнала.

EXPERIMENTAL SYSTEM FOR VISIBLE LIGHT COMMUNICATION CHANNEL ANALYSIS BASED ON SDR

A.M. SALAUYOU

Abstract. This paper presents an experimental system for the investigation of a visible light communication (VLC) channel, implemented using software-defined radio (SDR) platforms and an optoelectronic transceiver. The proposed setup provides signal generation, transmission, and reception with the ability to control optical channel parameters and propagation conditions. The system is designed for continuous acquisition of signal parameters, enabling the analysis of channel dynamics. This approach can be used to study the temporal structure of the channel and perform its further analysis.)

Keywords: visible light communication system; software-defined radio; OFDM; signal analysis; signal-to-noise ratio; temporal characteristics; channel dynamics.

Список литературы

1. Komine T., Nakagawa M. // IEEE Transactions on Consumer Electronics. 2004. Vol. 50. P. 100–107.
2. Abdul Wadud, Anas Basalamah // ICT Express. 2025. Vol. 11. P. 45–59.
3. Коган С. С. // Первая миля. 2023. № 7. С. 56–61.
4. López-Benítez M. // Physical Communication. 2018. Vol. 30. P. 154–166.
5. Chvojka P. et al. // IEEE Photonics Journal. 2015. Vol. 7. P. 1–10.
6. GNU Radio.
7. Armstrong J. // Journal of Lightwave Technology. 2009. Vol. 27. P. 189–204.
8. Abdullah M. J. et al. // Proc. of 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall). 2023. P. 1–2.
9. HackRF One: open source software defined radio platform. 2014.
10. LimeSDR: open source software defined radio platform. 2016.
11. Che M. // Telecom. 2025. Vol. 6. P. 31.
12. Costanzo A., Loscri V., Biagi M. // Journal of Lightwave Technology. 2021. Vol. PP. P. 1–1.
13. Kassem A., Darwazeh I. // Proc. of 2022 13th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP). 2022. P. 467–472.
14. Фотодиод ФД-263: характеристики, применение // Электронные компоненты и радиодетали. 2026.
15. Соловьев А. Н., Половения С. И. // Веснік сувязі. 2025. № 4. С. 58–64.

АВТОМАТИЗАЦИЯ ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЙ ДИАГРАММ В ПРОГРАММНЫЙ КОД

А.В. ЛОМАКО, К.А. ХАДЖИНОВА

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 06 апреля 2026

Аннотация. В статье предложен подход к автоматизации преобразования изображений диаграмм в программный код на основе конвейерной обработки. Разработана теоретическая модель, включающая распознавание графических объектов, построение графовой модели, формализацию в терминах конкретной нотации и генерацию кода через унифицированное промежуточное представление. Рассмотрены основные нотации моделирования (UML, BPMN 2.0, сети Петри, IDEF0, IDEF3, DFD, ERD) и их соответствие типам целевого кода. Описан алгоритм обработки изображений диаграмм с выделением этапов предобработки, детекции объектов, построения графа, классификации нотации и кодогенерации. Метод позволяет автоматизировать ручной этап переноса диаграмм в программную реализацию, обеспечивая согласованность модели и кода.

Ключевые слова: диаграммы, автоматизация, распознавание изображений, графовая модель, кодогенерация, UML, BPMN, промежуточное представление, компьютерное зрение, формализация нотаций.

Введение

В программных проектах диаграммы служат не только для визуализации, но и как формальный инструмент описания структуры, поведения и данных системы [1]. В процессе разработки диаграммы вручную переводятся в код, в результате чего графическое представление модели может не полностью совпадать с программной реализацией, занимает определенное время, растет вероятность ошибок и утрачивается исходный смысл модели [2].

Для решения этой проблемы предлагается автоматизация преобразования изображений диаграмм в программный код. В работе предложен последовательный конвейер преобразования изображения диаграммы в программный код, включающий распознавание, построение графовой модели, формализацию в нотацию и генерацию кода.

Обзор существующих решений

В CASE-средствах и системах модельно-ориентированной разработки используются формализованные модели, представленные в форматах UML/XMI, BPMN и во внутренних форматах редакторов [3]. Такие инструменты позволяют автоматически генерировать программный код, SQL-скрипты, конфигурации и другие артефакты, а также поддерживают двустороннюю синхронизацию между моделью и кодом, что сокращает ручной труд и повышает согласованность документации и реализации [4].

Однако большинство решений не рассматривают задачу работы с изображениями диаграмм, полученных в виде скриншотов, отсканированных документов или экспортированных графических файлов. В таком формате диаграмма лишена явной семантики, поэтому ее приходится вручную переносить в формализованную модель, и между изображением и программным кодом остается трудно автоматизируемый ручной этап [5].

Существующие подходы к автоматизации можно разделить на несколько направлений:

- генерация кода из формализованных UML/BPMN-моделей в CASE-средствах без работы с изображениями;
- распознавание графических примитивов и текста на изображениях диаграмм без дальнейшей формализации и генерации кода;
- системы генерации кода для отдельных нотаций (например, ERD, сети Петри) на основе жестких шаблонов без единой теоретической основы [6].

В литературе отсутствует единый подход для преобразования изображений диаграмм в программный код по широкому набору стандартных нотаций [7]. Предлагаемый в работе метод решает данную проблему за счет введения унифицированной теоретической модели и конвейерного алгоритма, позволяющего обрабатывать различные нотации на общей промежуточной и теоретической основе [8].

Классификация диаграмм и формальные основания

Для наглядного представления типов диаграмм и их групп по категориям приведена схема, показанная на рис. 1.



Рис. 1. Иерархия нотаций моделирования

Данная схема иллюстрирует группировку нотаций по типу описания (структура, поведение, процессы, сетевые модели) и по предметной области, показывая, что множество используемых нотаций естественно группируется в несколько типовых категорий. Это позволяет построить единую теоретическую модель преобразования, опирающуюся на единый графовый, и мета-модельный каркас, который унифицирует обработку диаграмм различных типов и позволяет использовать их в рамках одной и той же конвейерной схемы преобразования изображения в программный код.

В рамках предлагаемого подхода рассматриваются несколько основных нотаций, имеющих достаточную формализацию и допускающих прямую трансляцию в программные артефакты [9]. К ним относятся UML (включая диаграммы деятельности и классов), BPMN 2.0, сети Петри, DFD, ERD, IDEF0 и IDEF3 диаграммы [10]. Эти нотации охватывают структурные, поведенческие, процессные и сетевые модели систем, что позволяет использовать их в качестве единой теоретической основы для унифицированного описания и преобразования [11].

Соответствие нотаций, их формального представления и типов целевого кода подробно описано в табл. 1.

Из таблицы 1 видно, что каждая нотация описывается в виде структуры элементов и связей. Несмотря на различия в интерпретации и визуальном оформлении, все рассмотренные нотации можно представить в виде единого графового представления, в котором элементы диаграммы соответствуют вершинам, а зависимости между ними – ребрам этой структуре из теоретической модели, где элементы и связи выражаются через граф G и формальную модель M .

Табл. 1. Характеристики анализируемых графических моделей

Нотация	Тип диаграммы	Элементы	Логика	Код/результат
UML Activity	Поведенческая	Начальный узел, действия, переходы, решения (ромбы), слияния, конечный узел	$start \rightarrow action \rightarrow$ $if\ cond \rightarrow end$	Workflow-скрипт, бизнес-логика <pre>def process_order(order): if order.is_valid(): invoice = create_invoice(order) else: send_rejection(order)</pre>
UML Class	Структурная	Классы, свойства, операции, связи, наследование, композиция	$class\ Order; has$ $items[]; relates$ $Customer$	DTO, ORM-модели <pre>class Order: def __init__(self, customer_id: int): self.customer_id = customer_id self.items = []</pre>
BPMN 2.0	Процессная	События начала/окончания, задачи, развилки (исключающие/параллельные), потоки последовательности	$start \rightarrow task \rightarrow$ $exclusiveGateway \rightarrow end$	BPMN XML, orchestration-код <pre><process id="OrderProcess"> <startEvent id="start"/> <task id="checkPayment"/> <exclusiveGateway id="gateway"/> <endEvent id="end"/> </process></pre>
Сети Петри	Формальная сетевая	Места (круги), переходы (прямоугольники), дуги, жетоны (маркировка)	$if\ marking[p] \geq 1:$ $fire\ transition$	Модель верификации <pre>def fire_transition(marking, t): if all(marking[p] >= 1 for p in t.inputs): for p in t.inputs: marking[p] -= 1 for p in t.outputs: marking[p] += 1</pre>
IDEF0	Функционально-процессная	Функции (прямоугольники), входы (слева), выходы (справа), управление (сверху), механизмы (снизу)	$decompose$ $process \rightarrow$ $subproc1,$ $subproc2$	Спецификации процессов <pre>def process_data(input_data): validated = validate(input_data) transformed = transform(validated) return save(transformed)</pre>
IDEF3	Процессная	Процессы, шаги (единицы работ), объекты потоков, узлы связей, альтернативы	$process: step1 \rightarrow$ $step2 \rightarrow [alt]$ $step3$	Сценарии исполнения <pre>def etl_pipeline(): source = extract_from_db() if validate(source): transformed = transform(source) load_to_warehouse(transformed)</pre>
DFD	Поток данных	Процессы (круги), внешние сущности, хранилища данных, потоки данных	$input \rightarrow process1$ $\rightarrow datastore \rightarrow$ $output$	ETL-логика, архитектура данных <pre>def etl_pipeline(): raw_data = load_from_source() cleaned = clean_data(raw_data) save_to_database(cleaned)</pre>
ERD	Структурная модель данных	Сущности (прямоугольники), свойства (овалы), ключи, связи (ромбы), кратность	$Order\ 1:N$ $OrderItem;$ $Customer\ 1:1$ $Order$	SQL DDL, ORM-схема <pre>CREATE TABLE Orders (id SERIAL PRIMARY KEY, customer_id INT REFERENCES Customers(id), total DECIMAL(10,2));</pre>

Теоретическая модель преобразования

Для описания процесса преобразования изображения диаграммы I в программный код C вводится последовательность отображений, связывающих исходное изображение с конечным кодом через ряд промежуточных представлений. В рамках модели выделяются этапы: анализ

изображения и построение графовой модели, формализация в терминах конкретной нотации и построение промежуточного представления, из которого осуществляется генерация кода.

Общая схема преобразования задается формулой:

$$C = g \left(f_3 \left(f_2 \left(f_1 (I) \right) \right) \right), \quad (1)$$

где I – исходное изображение диаграммы; f_1 – функция распознавания объектов и построения графа; f_2 – функция формализации, сопоставляющая граф с соответствующей формальной моделью; f_3 – функция построения промежуточного представления; g – функция генерации кода из этого представления.

Формула (1) задает последовательность преобразований изображения в код в рамках единой конвейерной схемы. Исходное изображение рассматривается как двумерная матрица пикселей, что формально описывается в формуле:

$$I = \{p_{ij} \mid i = 1, \dots, m, j = 1, \dots, n\}, \quad (2)$$

где p_{ij} – значение яркости или цвета пикселя в позиции (i, j) ; m и n – высота и ширина изображения в пикселях.

Формула (2) задает базовое представление, с которым работают все этапы предобработки и распознавания [12]. Предобработка включает нормализацию изображения по размеру, контрасту и яркости, а также удаление шумов [13]. Результатом предобработки служит основа для построения множества распознанных объектов, задаваемого формулой:

$$O = \{o_k\}, o_k = (b_k, t_k, s_k), \quad (3)$$

где b_k – ограничивающий прямоугольник объекта на изображении; t_k – распознанный текст, ассоциированный с объектом; s_k – семантический тип объекта.

Формула (3) задает структуру данных для построения графа. На основе множества объектов O строится ориентированный граф, структура которого задается формулой:

$$G = (V, E), \quad (4)$$

где V – множество вершин, соответствующих элементам диаграммы; E – множество ребер, отражающих зависимости, потоки данных/управления, переходы или ассоциации между элементами.

Структура каждой вершины графа дополнительно описывается формулой:

$$v_i = (\text{type}_i, \text{attr}_i, \text{pos}_i), \quad (5)$$

где type_i – тип узла; attr_i – набор атрибутов, относящихся к элементу; pos_i – позиция элемента на изображении.

Формулы (4) и (5) определяют двойную роль графа G : с одной стороны, он хранит топологию связей, с другой – семантические атрибуты элементов диаграммы.

Формализация графа заключается в сопоставлении его элементов с элементами соответствующей формальной модели [14]. Это отображение задается функцией, которая на основе структуры графа и набора типов узлов и ребер определяет корректные элементы и правила модели, как указано в формуле:

$$M = f_2(G), \quad (6)$$

где M – формальная модель диаграммы в терминах выбранной нотации; f_2 – функция формализации, интерпретирующая граф G в рамках метамодели.

Формула (6) определяет модель M как интерпретацию графа G в выбранной нотации. Для унификации обработки различных нотаций вводится промежуточное представление, отделенное от этапов распознавания и генерации кода. Его структура в общем виде описывается формулой:

$$IR = (N, R, A), \quad (7)$$

где N – множество узлов, соответствующих элементам предметной области; R – множество отношений между ними; A – набор атрибутов, содержащих имена, типы, параметры, метки и кратность связей.

Формула (7) задает абстрактный уровень, на котором моделируются любые исследуемые нотации, независимо от их визуальной специфики.

Структура и типовые атрибуты компонентов IR подробно описаны в табл. 2.

Табл. 2. Структура промежуточного представления IR

Компонент	Описание	Примеры атрибутов
N – узлы	Элементы предметной области (сущности, задачи, места)	<i>name, type, subtype, description</i>
R – отношения	Связи, зависимости, потоки данных/управления	<i>source, target, type, condition, label</i>
A – атрибуты	Свойства, значения, метки	<i>name, value, cardinality, multiplicity</i>

Использование единого уровня IR позволяет унифицировать правила генерации кода для разных нотаций: функция g отображает IR в целевой код, не завися от конкретной исходной нотации, как указано в формуле:

$$C = g(IR), \quad (8)$$

где C – программа/скрипт/описание в целевом формате; g – функция генерации кода, зависящая от целевой нотации и языка программирования.

Формула (8) связывает промежуточное представление с конечным программным кодом.

Общая структура теоретической модели преобразования изображения диаграммы в программный код представлена на схематичной блок-диаграмме, показанной на рис. 2.



Рис. 2. Модель преобразования «изображение–код»

Теоретическая модель задает единый формальный каркас, в котором распознавание изображения, построение графа, формализация, промежуточное представление и кодогенерация становятся составными частями одной общей схемы, позволяющей унифицировать обработку диаграмм различных типов.

Алгоритм обработки изображения и генерации кода

Алгоритм реализует последовательность преобразований, заданных теоретической моделью, в виде конвейера, в котором изображение диаграммы проходит предобработку, распознавание объектов, построение графа, формализацию, построение промежуточного представления IR и генерацию целевого кода.

Общая структура алгоритма представлена на рис. 3 на схематичной блок-диаграмме.



Рис. 3. Схема алгоритма обработки изображения и генерации кода

Рис. 3 иллюстрирует последовательность этапов, в которой выход каждого блока передается на вход следующего: изображение $I \rightarrow$ предобработка $\rightarrow$ распознанные объекты $O \rightarrow$ граф $G \rightarrow$ формальная модель $M \rightarrow$ промежуточное представление $IR \rightarrow$ программа/скрипт C .

Основными промежуточными представлениями, через которые проходят данные, являются распознанные объекты O и граф G . Структура распознанных объектов задается формулой (3), а на их основе строится ориентированный граф G по формуле (4). Структура каждой вершины графа уточняется в формуле (5), что позволяет связать визуальное положение элемента с его семантикой. На следующем шаге граф анализируется для определения типа нотации, и формируется формальная модель M по формуле (6). Эта модель переносится в единое промежуточное представление IR по формуле (7), а из IR генерируется целевой код C по формуле (8).

Структура распознанных объектов и соответствующих вершин графа показана в табл. 3.

Табл. 3. Семантика вершин графовой модели

Элемент / тип объекта	Содержимое в O	Содержимое в вершине графа
Состояние / действие	b_k , текст, тип «state»	type, name, input, output, constraints
Задача / событие	b_k , текст, тип «task»	type, name, condition, triggers
Сущность (ERD)	b_k , текст, тип «entity»	type, name, attributes, cardinality
Переход (сети Петри)	b_k , текст, тип «transition»	type, name, firing_condition
Функция (DFD/IDEF)	b_k , текст, тип «function»	type, inputs, outputs, controls, mechanisms

Данная таблица показывает, как элементы множества O отображаются в вершины графа, сохраняя как визуальное положение, так и семантику, необходимые для последующей классификации нотации и формализации модели.

Алгоритм начинается с предобработки изображения: нормализации размера, контраста и яркости, а также фильтрации шумов. Далее с применением методов компьютерного зрения выделяются графические объекты диаграммы (прямоугольники, окружности, стрелки, текстовые блоки и др.). Для каждого объекта определяется ограничивающий прямоугольник и тип, а также выполняется распознавание текста по области этого прямоугольника [15]. На основе полученных объектов строится ориентированный граф, в котором вершины соответствуют элементам диаграммы, а ребра отражают зависимости, потоки данных/управления или ассоциации. Структура графа анализируется для определения типа нотации, после чего граф формализуется в соответствующую модель M , которая переносится в единое промежуточное представление IR . На завершающем этапе по этому представлению генерируется программный код для выбранной нотации и целевого языка программирования [16].

Алгоритм объединяет теоретическую модель и схему обработки, а также конкретную структуру данных (табл. 2 и 3), в единую схему, позволяющую преобразовывать изображение диаграммы в программный код и адаптировать результат под выбранную нотацию и целевой язык программирования.

Детальную логику алгоритма фиксирует следующий псевдокод:

```

def DiagramToCode(I):
    """
    Преобразование изображения диаграммы в программный код.
    Вход:
        I – изображение диаграммы, матрица пикселей.
    Выход:
        C – программа/скрипт/описание в целевом формате.
    """
    I_pre = preprocess_image(I)
    O_raw = detect_objects(I_pre)
    O = []
    for object_k in O_raw:
        box = bounding_box(object_k)
        text = ocr_single_box(box)
        type_ = classify_object_type(object_k.shape, object_k.position)
        entry = (box, text, type_)
        O.append(entry)
    G = build_graph_from_objects(O)
    notation = classify_notation(G)
    M = formalize_model(G, notation)
    IR = build_intermediate_representation(M)
    C = generate_code(IR, notation, target_language)
    return C

```

Ограничения и перспективы

Качество преобразования диаграммы в код зависит от точности и стабильности распознавания объектов, текста и построения графа. Ошибки, неоднозначная интерпретация примитивов и шумы могут привести к искажению структуры. Метод предполагает стандартный вид диаграммы и соответствие нотациям [17]. При отклонениях (нестандартное оформление, рисование от руки, смещение нотаций) возрастает риск ошибочного сопоставления элементов. IR и функция генерации кода оптимизированы под заданный набор языков и нотаций. Расширение на новые языки потребует доработки правил и шаблонов, увеличивая сопровождение.

В перспективе – интеграция продвинутых моделей компьютерного зрения и языковых моделей для повышения точности [18], расширение поддержки новых нотаций и языков, а также использование IR для анализа свойств моделей (корректность, тупики, избыточные связи) с совмещением кодогенерации и рефакторинга.

Заключение

Предложен подход преобразования изображения диаграммы в код на основе конвейерной обработки: от распознанных объектов и графовой модели до формализации, построения IR и генерации кода. Введена теоретическая модель преобразования через последовательность функций, формализованы структуры данных, графа и IR. Алгоритм реализует конвейер из этапов детекции, распознавания текста, построения графа, классификации нотации, формализации модели и кодогенерации. Архитектура конвейера показана на рисунке 3, структура закреплена в псевдокоде. Унифицированное IR обеспечивает гибкость и масштабируемость подхода для разных нотаций и языков.

Несмотря на ограничения, подход автоматизирует ручной этап преобразования диаграмм в код и может использоваться в инструментах анализа и генерации на основе скриншотов, сканов и экспортированных диаграмм.

AUTOMATION OF DIAGRAM IMAGE CONVERSION TO PROGRAM CODE

A.V. LOMAKO, K.A. KHADZHYNAYA

Abstract. The article proposes an approach to automating the conversion of diagram images into program code based on conveyor processing. A theoretical model has been developed, including recognition of graphical objects, construction of a graph model, formalization in terms of specific notation, and code generation through a unified intermediate representation. The main modeling notations (UML, BPMN 2.0, Petri nets, IDEF0, IDEF3, DFD, ERD) and their correspondence to target code types are considered. An algorithm for processing diagram images is described, highlighting the stages of preprocessing, object detection, graph construction, notation classification, and code generation. The method automates the manual stage of transferring diagrams to program implementation, ensuring model-code consistency.

Keywords: diagrams, automation, image recognition, graph model, code generation, UML, BPMN, intermediate representation, computer vision, notation formalization.

Список литературы

1. Fowler M. UML distilled: a brief guide to the standard object modeling language. 3-е изд. – Reading, MA : Addison-Wesley, 2003. – 209 с.
2. Unified Modeling Language (UML) superstructure. Version 2.5.1 / Object Management Group. – Needham, MA : OMG, 2017. – 768 с.
3. Business Process Model and Notation (BPMN). Version 2.0.2 / Object Management Group. – Needham, MA : OMG, 2018. – 557 с.
4. IEEE Standard Glossary of Software Engineering Terminology : IEEE Std 610.12-1990. – New York : IEEE, 1990. – 84 с.
5. Knuth D. E. The art of computer programming. Vol. 1 : Fundamental algorithms. 3rd ed. – Reading, MA : Addison-Wesley, 1997. – 650 с.
6. Reisig W. Understanding Petri nets: modeling techniques, analysis methods, case studies. – Berlin : Springer, 2013. – 222 с.
7. Goma H. Designing software product lines with UML: from use cases to pattern-based software architectures. – Reading, MA : Addison-Wesley, 2004. – 512 с.
8. Sommerville I. Software engineering. 10th ed. – Harlow : Pearson, 2016. – 775 с.
9. Gamma E. Design patterns: elements of reusable object-oriented software / E. Gamma, R. Helm, R. Johnson, J. Vlissides. – Reading, MA : Addison-Wesley, 1995. – 395 с.
10. Kruchten P. The rational unified process: an introduction. 3rd ed. – Reading, MA : Addison-Wesley, 2004. – 212 с.
11. Satyanarayanan M. Pervasive computing: vision and challenges // IEEE Personal Communications. – 2001. – Vol. 8, № 4. – P. 10–17.
12. Szeliski R. Computer vision: algorithms and applications. 2nd ed. – Springer, 2022. – 950 с.
13. Goodfellow I. Deep learning / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge, MA : MIT Press, 2016. – 775 с.
14. Куделев А. В. Автоматизация разработки программного обеспечения на основе моделей UML : дис. ... канд. техн. наук. – М. : МГТУ им. Н. Э. Баумана, 2023. – 180 с.
15. Иванов П. С. Распознавание структурных элементов блок-схем // Вестник МГТУ. Серия Приборостроение. – 2025. – № 1. – С. 23-34.
16. Петрова Е. А. Генерация SQL-скриптов из ER-диаграмм с использованием компьютерного зрения // Программные продукты и системы. – 2024. – № 2. – С. 56-67.
17. Смирнов А. Б. Моделирование бизнес-процессов в нотации BPMN 2.0. – СПб. : Питер, 2023. – 320 с.
18. Ли Ч. Компьютерное зрение для анализа технической документации / Ч. Ли, М. Ким. – М. : ДМК Пресс, 2024. – 456 с.

УДК 621.391

КОМПЛЕКСНАЯ СИСТЕМА УПРАВЛЕНИЯ БЕСПИЛОТНЫМ АВТОМОБИЛЕМ НА БАЗЕ ИОТ: ИНТЕГРАЦИЯ ЭНЕРГОПИТАНИЯ И ВИДЕОНАБЛЮДЕНИЯ

Д.Ю. ВОРОНОВИЧ, М.Ю. МАКСАК

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь**Поступила в редакцию 02 апреля 2026*

Аннотация. В статье рассматривается проблема разработки комплексной системы управления беспилотным автомобилем, основанной на технологии IoT. Актуальность исследования обусловлена необходимостью перехода от изолированных бортовых систем к интегрированной экосистеме, где подсистемы энергопитания и технического зрения функционируют согласованно. Предложена архитектура, объединяющая интеллектуальное управление энергопотреблением и системы видеонаблюдения с элементами компьютерного зрения. Особое внимание уделено взаимосвязи между нагрузкой на вычислительные узлы обработки видеопотока и стратегией распределения энергии, а также вопросам информационной безопасности передачи данных. Проанализированы подходы к интеграции транспортного средства с дорожной инфраструктурой и облачными платформами с учетом требований надежности и энергоэффективности. Результаты исследования демонстрируют, что совместная оптимизация алгоритмов управления питанием и видеоконтролем позволяет повысить автономность, безопасность и адаптивность беспилотных транспортных средств в условиях современной транспортной инфраструктуры.

Ключевые слова: беспилотный автомобиль, Интернет вещей (IoT), интеллектуальная система управления, энергоэффективность, видеонаблюдение, компьютерное зрение, архитектура системы, информационная безопасность.

Введение

Развитие автономных транспортных средств является одним из стратегических направлений современной транспортной инфраструктуры. В условиях роста требований к безопасности, энергоэффективности и интеллектуализации логистических процессов, беспилотные автомобили все чаще рассматриваются не как изолированные технические системы, а как интегрированные элементы единой интеллектуальной транспортной экосистемы на базе технологий IoT [1]. Особую роль в обеспечении функциональности таких систем играют подсистемы технического зрения и энергоменеджмента, которые традиционно проектируются раздельно, что снижает общую эффективность транспортного средства.

Системы видеонаблюдения и компьютерного зрения формируют основу для восприятия окружающей среды и принятия решений в реальном времени [2]. Однако обработка видеопотоков и работа нейросетевых алгоритмов требуют значительных вычислительных ресурсов, что напрямую влияет на энергопотребление бортовой сети. В то же время, интеллектуальные системы управления питанием, ориентированные на оптимизацию зарядки и распределения энергии, часто не учитывают динамическую нагрузку со стороны периферийных датчиков и вычислительных модулей [3]. Отсутствие координации между подсистемами энергопитания и видеоконтроля приводит к неоптимальному расходу заряда батареи и снижению автономности АТС в сложных условиях эксплуатации.

Дополнительную актуальность приобретают вопросы информационной безопасности и надежности передачи данных. В условиях, когда видеопотоки используются для критически важных решений, а система питания взаимодействует с внешней инфраструктурой, обеспечение

целостности и доступности информации становится неотъемлемой частью архитектуры системы. Это особенно важно в свете требований по защите персональных данных и технических регламентов безопасности колесных транспортных средств [4].

На сегодняшний день недостаточно исследованы подходы к построению распределенных IoT-ориентированных систем, сочетающих управление энергопотреблением и видеонаблюдением в едином контуре. Отсутствие комплексных решений, адаптированных к условиям интеграции с дорожной инфраструктурой и облачными платформами, определяет научную и практическую значимость настоящего исследования.

Целью работы является исследование архитектуры комплексной системы управления беспилотным автомобилем на базе IoT, обеспечивающей согласованное функционирование подсистем энергопитания и видеонаблюдения. Для достижения поставленной цели решаются задачи анализа существующих архитектур, разработки модели взаимодействия подсистем и оценки эффективности предложенных алгоритмов управления ресурсами.

Архитектура интегрированной системы управления

В рамках данного исследования предлагается многоуровневая архитектура комплексной системы управления беспилотным автомобилем, основанная на технологиях IoT. Основная идея заключается в объединении подсистем энергопитания и видеонаблюдения в единый информационно-управляющий контур, что позволяет оптимизировать распределение ресурсов в реальном времени.

На рис. 1 представлена обобщенная структурная схема предложенной системы.

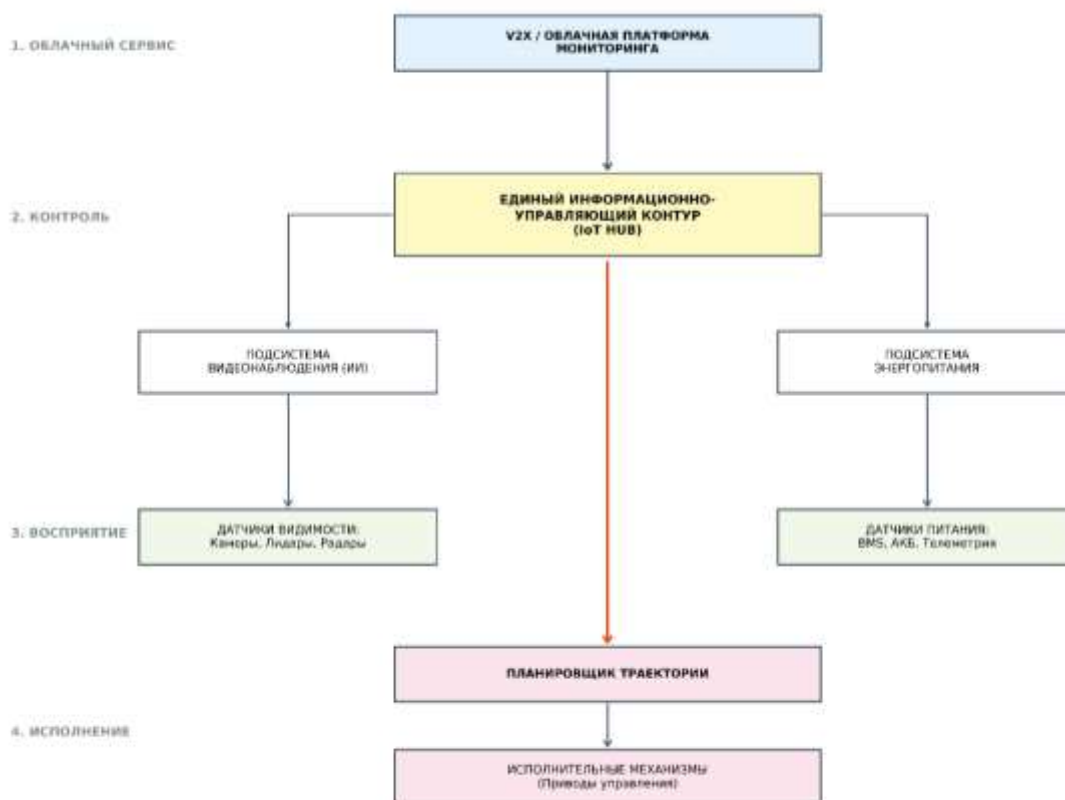


Рис. 1. Структурная схема интегрированной системы управления беспилотным автомобилем

На сенсорном уровне размещаются источники данных и исполняющие устройства. К ним относятся камеры технического зрения, датчики тока и напряжения бортовой сети, блок управления батареей и вычислительные модули. Важной особенностью является наличие единого IoT-шлюза, который агрегирует данные от разнородных источников [5].

Сетевой уровень обеспечивает передачу данных между компонентами системы. Для связи внутри автомобиля используется шина Controller Area Network, обеспечивающая

детерминированную передачу критических данных, таких как показания датчиков напряжения. Для передачи видеопотоков и телеметрии во внешнюю инфраструктуру применяются беспроводные протоколы и легковесные протоколы прикладного уровня, такие как MQTT.

Уровень приложений реализует логику управления. Здесь функционируют алгоритмы компьютерного зрения (детекция объектов, построение маршрута) и алгоритмы энергоменеджмента. В отличие от традиционных подходов, где эти подсистемы работают изолированно, в предложенной архитектуре реализован обратный канал влияния: данные о вычислительной нагрузке систем зрения передаются в модуль управления питанием.

Взаимодействие подсистем организовано следующим образом. Модуль видеонаблюдения формирует запрос на вычислительные ресурсы в зависимости от сложности дорожной сцены. Эта информация передается в интеллектуальный контроллер питания, который корректирует режимы работы батареи и распределение энергии между двигателем и бортовой электроникой. Данные о состоянии заряда батареи могут влиять на работу видеосистемы, например, снижая частоту кадров или разрешение потока при критическом уровне энергии для продления автономности [3]. Такая интеграция позволяет реализовать адаптивные сценарии работы. Например, при обнаружении препятствия системой зрения приоритет энергии временно отдается вычислительному блоку для экстренного расчета траектории, а при движении по трассе с хорошим освещением нагрузка на камеры снижается в пользу экономии заряда.

Таким образом, предложенная архитектура обеспечивает не только сбор и обработку данных, но и интеллектуальное управление ресурсами автомобиля на основе комплексного анализа состояния как энергетической системы, так и системы восприятия окружающей среды.

Модель взаимодействия подсистем и алгоритм управления ресурсами

Для описания зависимости энергопотребления от нагрузки систем технического зрения введем целевую функцию оптимизации. Пусть E_{total} – общее доступное энергопотребление на текущем участке маршрута, E_{drive} – энергия, необходимая для движения, E_{comp} – энергия на вычислительные задачи. Тогда ограничение имеет вид

$$E_{comp} \leq E_{total} - E_{drive} - E_{reserve}, \quad (1)$$

где $E_{reserve}$ – резервный запас энергии для нештатных ситуаций.

Энергопотребление вычислительного модуля E_{comp} напрямую зависит от параметров видеопотока. Его можно представить как функцию от частоты кадров FPS , разрешения видеопотока Res и сложности алгоритма детекции Alg

$$E_{comp} = k_1 \cdot FPS \cdot Res + k_2 \cdot C_{alg}, \quad (2)$$

где k_1, k_2 – коэффициенты, определяемые архитектурой вычислительной платформы.

Сложность алгоритма C_{alg} варьируется в зависимости от дорожной сцены. Например, движение по трассе в дневное время требует меньше вычислительных ресурсов, чем движение в плотном городском потоке ночью. Система видеонаблюдения оценивает сложность сцены S_{scene} и передает эту метку в контроллер питания. На основе полученных данных реализуется алгоритм динамического распределения ресурсов [5]. Схема алгоритма представлена на рис. 2.

Алгоритм функционирует в цикле реального времени, начиная с мониторинга состояния системы, в ходе которого осуществляется сбор данных о State of Charge (SoC) батареи, текущей нагрузке на процессор и сложности дорожной сцены. На этапе оценки рисков при значении $SoC < 20\%$ система переходит в режим энергосбережения. Принятие решения основывается на анализе текущего состояния: при низком заряде батареи снижается частота кадров видеопотока ($FPS \downarrow$) или разрешение ($Res \downarrow$), что, согласно формуле (2), уменьшает E_{comp} , тогда как при обнаружении критических объектов (пешеходы, препятствия) приоритет отдается подсистеме зрения, даже ценой временного увеличения энергопотребления. На завершающем этапе

исполнения контроллер питания перераспределяет мощность между тяговой батареей и бортовой электроникой [6].

Таким образом, предложенная модель позволяет адаптивно управлять ресурсами автомобиля, обеспечивая баланс между автономностью и безопасностью движения. Реализация данного алгоритма была проверена в среде имитационного моделирования, результаты которого представлены в следующем разделе [7].

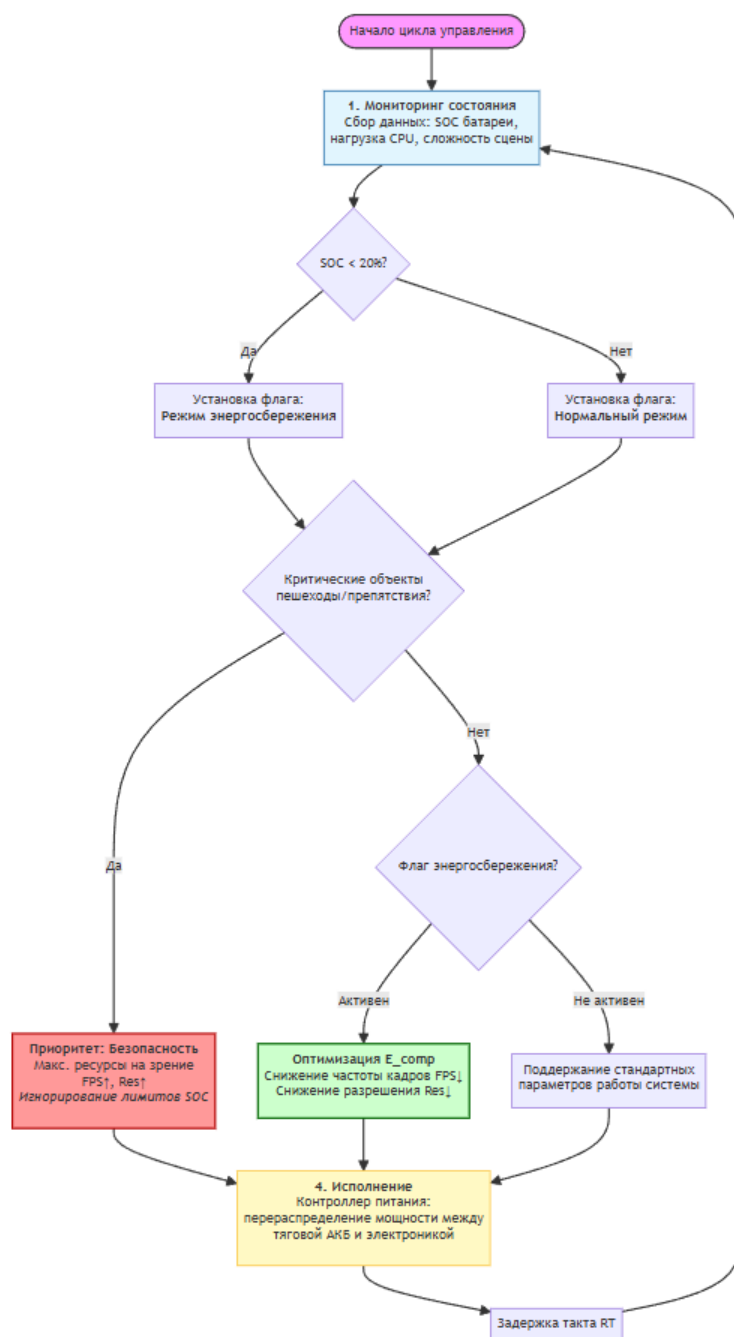


Рис. 2. Блок-схема алгоритма управления ресурсами

Результаты имитационного моделирования

Для оценки эффективности предложенной архитектуры и алгоритма управления ресурсами было проведено имитационное моделирование. Моделировался сценарий движения беспилотного автомобиля по городскому маршруту протяженностью 10 км с переменной дорожной обстановкой (открытая трасса, плотный городской трафик, ночное время).

В качестве метрик эффективности использовались:

- общее энергопотребление системы E_{total} ;
- точность детекции объектов Acc_{det} ;
- время автономной работы T_{auto} .

Сравнивались два подхода: базовый (независимая работа подсистем видео и питания) и предложенный (интегрированное управление с адаптивной настройкой параметров видеопотока).

Табл. 1. Сравнение эффективности подходов

Метрика	Базовый подход	Предложенный подход	Улучшение
E_{total} , Вт·ч	145,2	118,7	18,3 %
Acc_{det} , %	94,1	93,8	–0,3 %
T_{auto} , мин	87	106	+21,8 %

Как видно из таблицы, предложенный подход позволяет сократить общее энергопотребление на 18,3 % при незначительном снижении точности детекции, что находится в пределах допустимой погрешности. Прирост времени автономной работы составляет почти 22 %, что является существенным преимуществом для практического применения [8].

На рис. 3 показана динамика потребления энергии в зависимости от сложности дорожной сцены. В моменты высокой нагрузки предложенный алгоритм временно увеличивает выделение ресурсов на подсистему зрения, что обеспечивает безопасность. В простых сценариях система переходит в энергосберегающий режим, снижая частоту кадров и разрешение без потери критически важной информации.

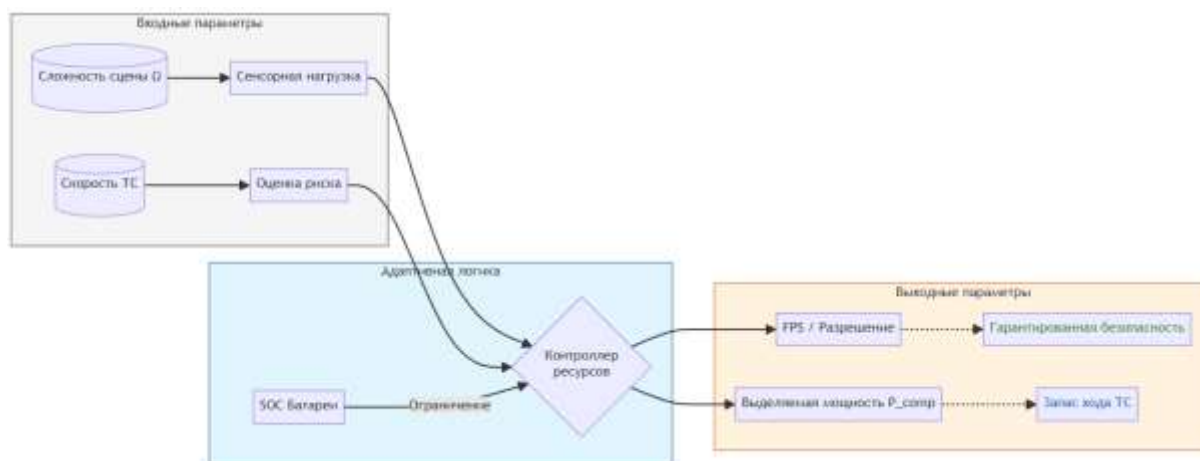


Рис. 3. Динамика энергопотребления в зависимости от сложности сцены

Полученные результаты подтверждают работоспособность предложенной архитектуры и эффективность алгоритма адаптивного управления ресурсами. Интеграция подсистем видеонаблюдения и энергопитания позволяет достичь баланса между автономностью и безопасностью, что является ключевым требованием для беспилотных транспортных средств.

Заключение

В работе решена задача интеграции подсистем энергопитания и видеонаблюдения беспилотного автомобиля в рамках единой IoT-архитектуры. Предложен алгоритм адаптивного управления ресурсами, динамически корректирующий параметры видеопотока в зависимости от сложности сцены и уровня заряда батареи. Результаты имитационного моделирования подтвердили эффективность разработанного подхода: применение алгоритма позволяет сократить общее энергопотребление на 18,3 % при незначительном снижении точности детекции

объектов (0,3 п.п.), что находится в пределах допустимой погрешности. Прирост времени автономной работы составляет почти 22 %, что является существенным преимуществом для практического применения беспилотных платформ в условиях городской и сельской местности.

Практическая значимость результатов заключается в возможности их использования при разработке отечественных беспилотных платформ, а также при модернизации существующих систем управления автономным транспортом. Учетные требования информационной безопасности каналов передачи данных обеспечивают защиту от несанкционированного вмешательства в процесс управления. Перспективы дальнейших исследований связаны с аппаратной реализацией предложенных алгоритмов на базе встраиваемых вычислительных платформ и проведением натурных испытаний в реальных дорожных условиях, включая адаптацию для работы в сетях 5G/6G с ультранизкой задержкой.

INTEGRATED IOT-BASED DRIVERLESS VEHICLE CONTROL SYSTEM: POWER SUPPLY AND VIDEO SURVEILLANCE INTEGRATION

D.Y. VORONOVICH, M.Y. MAKSAK

Abstract. An algorithm for searching for single-pixel extremes of halftone images based on centrally symmetric scanning is proposed. It is shown that the algorithm works much faster than the best known algorithms for detecting key points of images.

Keywords: driverless car, Internet of Things (IoT), intelligent control system, energy efficiency, video surveillance, computer vision, system architecture, information security.

Список литературы

1. Вишняков В. А., Максак М. Ю. Концепция, модель, примеры систем Интернета вещей для эффективного электроснабжения // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. / Белорус. гос. ун-т информатики и радиоэлектроники. – Минск: БГУИР, 2025. – С. 37–41.
2. Сиротко С. И., Ма Ц., Зорько М. И. [и др.]. Модели одновременной локализации и построения карты открытого пространства БЛА с помощью многокамерной видеосистемы // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. – Минск: БГУИР, 2025. – С. 7–14.
3. Хаджинова Н. В., Савоневская М. О., Столярова В. В. [и др.]. Алгоритмы мультиагентного доступа к каналам мобильной связи (5G, 6G) для минимизации задержек в условиях Интернета вещей // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. – Минск: БГУИР, 2025. – С. 67–72
4. Крагель В. А., Романюк М. В. Перспективы использования мобильных сетей нового поколения в Интернете вещей // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. – Минск: БГУИР, 2025. – С. 183–187
5. Бегляк Е. В., Марцинкевич В. А. Современные протоколы аутентификации и авторизации в беспроводных сетях // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. – Минск: БГУИР, 2025. – С. 168–175.
6. Thrun S., Burgard W., Fox D. Probabilistic Robotics. – Cambridge : MIT Press, 2005. – 680 p.
7. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement // arXiv preprint arXiv:1804.02767. – 2018.
8. Chen S., Hu Y., Shi Y. [et al.]. Vehicle-to-Everything (V2X) Communication: A Survey on Architecture, Security, and Applications // IEEE Communications Surveys & Tutorials. – 2023. – Vol. 25, № 2. – P. 1234–1267.

ОБЗОР И АНАЛИЗ МИРОВЫХ ТЕНДЕНЦИЙ ПОСТРОЕНИЯ СИСТЕМ ЗАЩИТЫ МЕСТ ПРОКЛАДКИ КАБЕЛЕЙ СВЯЗИ ОТ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА

Е.А. МАСЕЙЧИК, А.А.ПИЛЮШКО

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 02 апреля 2026

Аннотация. Статья посвящена анализу мировых тенденций построения систем защиты мест прокладки кабелей связи от несанкционированного доступа с учетом современных вызовов, касающихся безопасности телекоммуникационной инфраструктуры Республики Беларусь. Актуальность данной тематики обусловлена ростом уязвимости линейно-кабельных сооружений к физическим воздействиям и несанкционированному доступу, способным нарушить функционирование объектов критической инфраструктуры. Рассмотрены шесть направлений развития систем защиты: инженерные барьерные средства, видеонаблюдение, вибрационные и акустические датчики, распределенные волоконно-оптические системы, интеграция с IoT и применение методов искусственного интеллекта. Установлено, что большинство решений ориентированы на фиксацию уже произошедшего воздействия, что ограничивает возможности раннего предупреждения инцидентов. Наиболее перспективным направлением признано развитие распределенных волоконно-оптических систем мониторинга с интеллектуальной обработкой сигналов. Обоснована целесообразность перехода к интегрированным системам мониторинга протяженных объектов связи с элементами прогнозной аналитики.

Ключевые слова: методы и системы защиты мест прокладки кабелей связи от несанкционированного доступа, волоконно-оптические распределительные системы мониторинга.

Введение

Анализ современной научно-технической литературы [1–14], посвященной анализу мировых тенденций построения систем защиты (охраны, оповещения и т. п.) как отдельных (компактных) объектов, так и обширных территорий, в том числе мест прокладки кабелей связи (линейно-кабельных сооружений) от природно-техногенных причин и несанкционированного доступа показывает, что в условиях резко изменяющейся климатической, а также внутри- и внешнеполитической обстановки в Республике Беларусь в последние 5–6 лет актуальность разработки и внедрения таких систем многократно возросла.

При этом существующие решения, как правило, ориентированы на фиксацию уже совершенного воздействия, что снижает эффективность предотвращения инцидентов. В этих условиях особое значение приобретает разработка и внедрение систем раннего обнаружения вторжений на протяженных объектах связи. Цель статьи – провести классификацию и сравнительный анализ современных направлений развития методов и систем защиты кабельной инфраструктуры и определить наиболее перспективное решение для дальнейшего совершенствования.

Обзор и анализ мировых тенденций построения систем защиты мест прокладки кабелей связи

Современные подходы к защите мест прокладки кабелей связи формируются под влиянием требований обеспечения устойчивости телекоммуникационной инфраструктуры, а также необходимости минимизации рисков физического воздействия и несанкционированного

доступа. Международная практика демонстрирует переход от изолированных инженерных решений к многоуровневым системам мониторинга и реагирования, интегрирующим физические, технические и информационные средства защиты [9, 10]. На основе анализа научных публикаций и отраслевых отчетов выделяются шесть ключевых технологий развития систем защиты протяженных объектов связи.

Физическая инженерная защита (барьерные и пассивные средства)

Физическая инженерная защита представляет собой традиционный и наиболее распространенный способ обеспечения сохранности линейно-кабельных сооружений. Она включает прокладку кабелей в кабельной канализации, использование бронированных оболочек, металлических и железобетонных коробов, запираемых люков, антивандальных шкафов, а также создание охранных зон. Данные меры направлены на ограничение прямого физического доступа к кабелю и снижение вероятности его механического повреждения.

Практика проектирования систем охраны периметра показывает, что инженерные барьерные средства обладают высокой надежностью при условии соблюдения нормативных требований и регулярного технического обслуживания [7]. Их достоинствами являются автономность функционирования, отсутствие зависимости от электропитания и сетевой инфраструктуры, а также сравнительно низкие эксплуатационные расходы. Такие решения устойчивы к отказам и не требуют сложной калибровки.

Вместе с тем физическая защита не обеспечивает интеллектуального контроля состояния объекта. Она не позволяет выявить подготовительные действия нарушителя до момента фактического вскрытия или повреждения конструкции. Анализ принципов построения систем охраны периметра показывает, что барьерные средства носят пассивный характер и выполняют функцию сдерживания, а не раннего обнаружения угрозы [6]. При отсутствии средств мониторинга фиксируется лишь факт нарушения целостности защитного элемента, что затрудняет оперативное реагирование.

Дополнительным ограничением является невозможность масштабного контроля протяженных участков без значительных капитальных затрат. В условиях увеличения длин трасс и пространственной распределенности их инфраструктуры эффективность исключительно инженерных решений снижается. Таким образом, физическая защита остается базовым элементом такой системы безопасности, но не обеспечивает раннего детектирования вторжения.

Системы видеонаблюдения и периметрального контроля

Системы видеонаблюдения (CCTV), тепловизионные комплексы и интеллектуальная видеоаналитика широко применяются для охраны узлов связи, распределительных шкафов, кабельных колодцев и иных точек повышенного риска. Они позволяют осуществлять визуальный контроль территории, фиксировать попытки проникновения и проводить постфактум анализ событий.

Оценка эффективности периметральных систем демонстрирует, что визуальная верификация существенно повышает достоверность выявления инцидентов и снижает вероятность ложных тревог [8]. Интеграция с аналитическими алгоритмами обеспечивает автоматическое распознавание движения, пересечения заданных зон и подозрительных действий. Зарубежный опыт применения систем тревожного оповещения подтверждает рост их роли в обеспечении территориальной безопасности объектов критической инфраструктуры [4].

Однако видеонаблюдение имеет ряд существенных ограничений. Его эффективность зависит от погодных условий, освещенности и отсутствия физических препятствий. Наличие «слепых зон» снижает полноту контроля, особенно на протяженных трассах вне населенных пунктов. Кроме того, большинство видеоаналитических систем фиксирует уже происходящее событие, не обеспечивая прогнозирования или раннего предупреждения о потенциальной угрозе [6].

Для протяженных линейных объектов организация сплошного видеоконтроля экономически и технически затруднительна. Это ограничивает применение данной технологии преимущественно узловыми участками инфраструктуры.

Датчики вибрации и акустического контроля

Системы на основе вибрационных и акустических датчиков предназначены для регистрации механических воздействий на ограждения, грунт или конструкции, вблизи которых проложен кабель. Принципы их работы основаны на фиксации колебаний, возникающих при попытке подкопа, вскрытия люков или разрушения защитных элементов.

Тактико-технический анализ физических принципов обнаружения нарушителя показывает, что данные средства способны выявлять ранние стадии вторжения при корректной настройке чувствительности датчиков [6]. Применение распределенных сенсорных модулей позволяет организовать контроль протяженных участков с формированием сигнала тревоги при превышении пороговых значений [3].

К преимуществам систем, построенных по данной технологии, относятся их относительно высокая чувствительность, возможность автономной работы и адаптация к различным типам ограждений и грунтов. Вместе с тем характерной проблемой является значительное количество ложных срабатываний, вызванных воздействием ветра, транспорта, животных или климатических факторов. Для снижения числа ошибочных тревог требуется сложная фильтрация сигналов и настройка параметров, что повышает требования к квалификации обслуживающего персонала [5].

Таким образом, вибрационные и акустические датчики обеспечивают более раннее обнаружение по сравнению с пассивными средствами, однако их эффективность существенно зависит от корректности алгоритмов обработки данных и условий эксплуатации.

Волоконно-оптические распределенные системы мониторинга

Одним из наиболее динамично развивающихся направлений является использование распределенных волоконно-оптических систем мониторинга (DAS, DTS, DSS), в которых само оптическое волокно выполняет функцию чувствительного элемента. Принцип работы таких систем основан на анализе изменений параметров отраженного сигнала при механическом, температурном или акустическом воздействии на кабель.

Распределенные системы акустического мониторинга обеспечивают непрерывный контроль протяженных участков с высокой точностью локализации события [11]. Технологии DAS позволяют фиксировать вибрации вдоль всей длины кабеля с определением координаты воздействия, что особенно актуально для магистральных линий связи. Международные рекомендации по защите подводных кабелей подчеркивают необходимость применения мониторинговых решений для предупреждения преднамеренных и непреднамеренных повреждений [12].

К достоинствам данной тенденции относятся:

- непрерывный контроль на десятках километров без установки множества отдельных датчиков;
- высокая чувствительность к различным типам воздействий;
- возможность интеграции с аналитическими системами обработки сигналов.

Основными ограничениями являются высокая стоимость внедрения, сложность калибровки и необходимость применения специализированных алгоритмов интерпретации данных. Дополнительно отмечается потребность в защите информационных потоков системы мониторинга в соответствии с требованиями безопасности телекоммуникационной инфраструктуры [9, 14].

Несмотря на указанные сложности, распределенные волоконно-оптические системы демонстрируют потенциал обеспечения раннего выявления механических воздействий еще до фактического повреждения кабеля.

Интеграция с IoT и системами кибербезопасности

Современные системы защиты все чаще строятся по принципу интеграции датчиков различного типа в единую информационную платформу с передачей данных в централизованные центры мониторинга. Использование IoT-технологий позволяет организовать удаленный контроль состояния объектов, формировать события в реальном времени и осуществлять корреляцию сигналов от различных источников.

Аналитические материалы в области защиты телекоммуникационной инфраструктуры указывают на необходимость комплексного учета физических и киберугроз при построении систем безопасности [10, 14]. Интеграция с системами управления доступом, видеонаблюдением и распределенными датчиками создает основу для построения многоуровневой архитектуры защиты.

Преимуществами данной тенденции являются масштабируемость, централизованный контроль, автоматизация обработки событий и возможность накопления статистических данных для анализа рисков. Вместе с тем возрастает зависимость от сетевой инфраструктуры и устойчивости каналов передачи данных. Возникают дополнительные киберриски, связанные с возможностью вмешательства в систему мониторинга или подмены данных [9].

Таким образом, интеграция IoT расширяет функциональные возможности систем защиты, но требует соблюдения требований информационной безопасности и резервирования каналов связи.

Использование искусственного интеллекта и машинного обучения для раннего обнаружения

Развитие распределенных сенсорных систем и накопление больших массивов данных о состоянии протяженных объектов обусловили формирование отдельной тенденции – применение методов искусственного интеллекта и машинного обучения для обработки сигналов и раннего выявления угроз. В отличие от традиционных пороговых алгоритмов, интеллектуальные модели способны учитывать совокупность признаков, анализировать временные зависимости и выявлять аномальные паттерны поведения.

В распределенных волоконно-оптических системах мониторинга применение алгоритмов классификации и глубинного обучения позволяет дифференцировать типы воздействий: движение транспорта, шаги человека, работы строительной техники, попытки подкопа или вскрытия защитных конструкций [11]. Это снижает уровень ложных тревог и повышает точность идентификации события. Аналогичные подходы реализуются при обработке сигналов вибрационных датчиков и систем видеонаблюдения, где используются алгоритмы распознавания образов и поведенческого анализа.

Интеллектуальная обработка данных позволяет перейти от фиксации факта воздействия к прогнозированию вероятности развития инцидента. При накоплении статистики возможна реализация предиктивной аналитики, основанной на выявлении повторяющихся сценариев и корреляции сигналов от различных подсистем. В контексте обеспечения безопасности телекоммуникационной инфраструктуры такие решения соответствуют международным требованиям комплексного подхода к защите критических сервисов [9, 10].

К достоинствам данной тенденции относятся:

- снижение количества ложных срабатываний за счет адаптивной фильтрации сигналов;
- повышение точности локализации и классификации событий;
- возможность раннего выявления аномалий, предшествующих фактическому повреждению кабеля;
- интеграция в централизованные системы мониторинга.

Основными ограничениями являются необходимость формирования обучающих выборок, зависимость качества работы алгоритмов от полноты и репрезентативности данных, а также потребность в вычислительных ресурсах и квалифицированном сопровождении. Дополнительным фактором риска выступает уязвимость интеллектуальных систем к кибервоздействиям и попыткам подмены входных данных, что требует соблюдения требований информационной безопасности [14].

Таким образом, использование методов искусственного интеллекта формирует основу для построения систем раннего обнаружения вторжений, однако требует технологической зрелости инфраструктуры и интеграции с распределенными сенсорными платформами.

Заключение

Проведенный анализ шести основных направлений развития систем защиты протяженных объектов связи показывает, что каждое из них решает определенный класс задач, однако их эффективность существенно различается в части обеспечения раннего выявления угроз.

Физическая инженерная защита обеспечивает базовый уровень устойчивости инфраструктуры и снижает вероятность случайного повреждения, но не формирует механизмов оперативного мониторинга [6, 7]. Она носит преимущественно сдерживающий характер и не позволяет фиксировать подготовительные действия нарушителя.

Системы видеонаблюдения и периметрального контроля повышают достоверность выявления инцидентов за счет визуальной верификации, однако остаются зависимыми от внешних условий и не обеспечивают непрерывного контроля протяженных трасс [4, 8]. Их реактивный характер проявляется в фиксации уже происходящего события, без возможности анализа скрытых воздействий на грунт или подземные коммуникации.

Вибрационные и акустические датчики обеспечивают более раннее реагирование на механические воздействия по сравнению с пассивными средствами. Тем не менее высокая чувствительность к внешним факторам приводит к значительному количеству ложных тревог, что требует применения сложных алгоритмов фильтрации [5, 6]. Без интеллектуальной обработки их эффективность ограничена.

Распределенные волоконно-оптические системы мониторинга обладают принципиально иными характеристиками. Они обеспечивают непрерывный контроль всей длины кабеля, точную локализацию воздействия и возможность интеграции с аналитическими модулями [11]. Международные рекомендации по защите кабельной инфраструктуры подчеркивают необходимость применения мониторинговых решений для предупреждения инцидентов, а не только их фиксации [12]. При этом высокая стоимость и сложность внедрения ограничивают масштаб их применения без дополнительного экономического обоснования.

Интеграция систем защиты с IoT-платформами и средствами кибербезопасности позволяет объединить разнородные источники данных в единую архитектуру мониторинга [9, 10, 14]. Однако при отсутствии интеллектуальной аналитики такая интеграция не обеспечивает качественного перехода к раннему детектированию, а лишь повышает управляемость существующих средств.

Использование искусственного интеллекта и машинного обучения формирует условия для перехода от реактивной к проактивной модели защиты. Интеллектуальная корреляция сигналов различных датчиков позволяет выявлять аномалии до момента физического повреждения кабеля. В совокупности с распределенными сенсорными технологиями данный подход обеспечивает наибольший потенциал раннего обнаружения вторжения.

Обобщенный сравнительный анализ показывает, что большинство традиционных систем ориентированы на фиксацию уже произошедшего воздействия. Их общий недостаток заключается в отсутствии гарантированного механизма раннего предупреждения о готовящемся или начавшемся вторжении на территорию объекта либо на конкретный участок периметра. Наиболее перспективным направлением является интеграция распределенных волоконно-оптических систем мониторинга с интеллектуальной аналитикой, обеспечивающая непрерывный контроль, точную локализацию и снижение количества ложных тревог при соблюдении требований информационной безопасности [9, 11]. Поэтому основные усилия целесообразно сосредоточить на разработке именно таких систем с применением методов искусственного интеллекта и машинного обучения. Данный подход позволит обеспечить непрерывный контроль протяженных участков, точную локализацию воздействия и возможность дифференциации типов событий [11].

Интеграция распределенных сенсорных технологий с интеллектуальной аналитикой позволит реализовать переход от реактивной модели реагирования к проактивной, ориентированной на раннее выявление аномалий и предупреждение инцидентов. Такая архитектура соответствует международным требованиям комплексной защиты телекоммуникационной инфраструктуры и учета физических и киберугроз [9, 10, 14]. Дополнительным преимуществом является возможность централизованного мониторинга и накопления статистических данных для совершенствования алгоритмов обнаружения.

REVIEW AND ANALYSIS OF GLOBAL TRENDS IN BUILDING SYSTEMS TO PROTECT COMMUNICATION CABLE INSTALLATION SITES FROM UNAUTHORIZED ACCESS

E.A. MASEYCHIK, A.A. PILYUSHKO

Abstract. The article is devoted to the analysis of global trends in the construction of systems for protecting communication cable installations from unauthorized access, taking into account modern challenges related to the security of the telecommunications infrastructure of the Republic of Belarus. The relevance of this topic is due to the increasing vulnerability of linear cable structures to physical impacts and unauthorized access that can disrupt the functioning of critical infrastructure facilities.

Six areas of protection systems development are considered: engineering barriers, video surveillance, vibration and acoustic sensors, distributed fiber-optic systems, integration with IoT and the use of artificial intelligence methods.critical infrastructure.

It has been established that most solutions are focused on fixing the impact that has already occurred, which limits the possibilities of early warning of incidents. The development of distributed fiber-optic monitoring systems with intelligent signal processing is recognized as the most promising direction. The expediency of switching to integrated monitoring systems for extended communication facilities with elements of predictive analytics is substantiated.

Keywords: methods and systems for protecting communication cable installation sites from unauthorized access, fiber-optic distribution monitoring systems.

Список литературы

1. Берсенева, В. П. Комплекс охраны периметров «Мурена-КС» как средство защиты объектов / В. П. Берсенева // Охранные системы и технологии. – 2021. – URL: <https://cyberleninka.ru/article/n/kompleks-ohrany-perimetrov-murena-ks/pdf>
2. Здоровцов, А. Г. Способы применения дистанционно управляемых средств охранной сигнализации для обнаружения проникновений / А. Г. Здоровцов // Безопасность и связь. – 2022. – № 12. – С. 92–96. – URL: <https://cyberleninka.ru/article/n/sposoby-primeneniya-distantsionno-upravlyaemyh-sredstv-ohrannoy-signalizatsii-dlya-obnaruzheniya-proniknoveniy-na-ohranyaemye/pdf>.
3. Костюк, А. И. Мониторинг перемещения охранных модулей в системе охраны периметра / А. И. Костюк // Автоматизация и безопасность. – 2022. – URL: <https://cyberleninka.ru/article/n/monitoring-peremescheniya-ohrannyh-moduley-v-sisteme-ohrany-perimetra>
4. Курбанбаев, Д. Ж. Зарубежный опыт современного состояния применения систем тревожного оповещения в обеспечении территориальной безопасности / Д. Ж. Курбанбаев // Информационные технологии и безопасность. – 2023. – URL: <https://cyberleninka.ru/article/n/zarubezhnyy-opyt-sovremennogo-sostoyaniya-primeneniya-sistem-trevozhnogo-opovescheniya-v-sfere-obespecheniya-territorialnoy/pdf>.
5. Магомедов, Ш. Г. Построение моделей описания рисков охранных действий по защите внешних периметров организаций / Ш. Г. Магомедов // Информационная безопасность и риски. – 2017. – URL: <https://cyberleninka.ru/article/n/postroenie-modeley-opisaniya-riskov-ohrannyh-deystviy-po-zaschite-vneshnih-perimetrov-organizatsii>
6. Николаев, В. А. Тактико-технический анализ физических принципов обнаружения нарушителя в системах охраны периметра / В. А. Николаев // Системы безопасности. – 2017. – URL: <https://cyberleninka.ru/article/n/taktiko-tehnicheskii-analiz-fizicheskikh-printsipov-obnaruzheniya-narushitelya/pdf>
7. Селищев, В. А., Чечуга, О. В. Выбор системы охраны периметра / В. А. Селищев, О. В. Чечуга // Известия Тульского государственного университета. Технические науки. – 2010. – URL: <https://cyberleninka.ru/article/n/vybor-sistemy-ohrany-perimetra>
8. Щербакова, И. В. Анализ эффективности систем охраны периметра объектов / И. В. Щербакова // Информационные технологии безопасности. – 2016. – URL: <https://cyberleninka.ru/article/n/analiz-effektivnosti-sistem-ohrany-perimetra-obektov>
9. CIP Association. Legal and Regulatory Aspects Relating to the Physical Security of the Telecommunications Infrastructure Used for Critical Communication Services : report. – CIP Association, 2025. – URL: <https://www.cip-association.org/legal-and-regulatory-aspects-relating-to-the-physical-security-of-the-telecommunications-infrastructure-used-for-critical-communication-services/>
10. European Union Agency for Cybersecurity (ENISA). ENISA Threat Landscape for 5G Networks : report. – Athens : ENISA, 2023. – URL: <https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks>.

11. Hartog, A. H. A distributed optical fibre sensor for the detection and classification of disturbances along a fibre link / A. H. Hartog // Journal of Lightwave Technology. – 2020. – Vol. 38 (16), P. 4424–4436. – URL: <https://ieeexplore.ieee.org/document/9098927>
12. International Cable Protection Committee (ICPC). Recommendations for the Protection of Submarine Cables : report. – London : ICPC, 2024. – URL: <https://www.iscpc.org/publications/recommendations>
13. Kwon, Y. Distributed Fiber-Optic Sensing Technologies for Structural Health Monitoring / Y. Kwon // Photonics. – 2021. – Vol. 8 (1), P. 15. – URL: <https://www.mdpi.com/2304-6732/8/1/15>
14. National Institute of Standards and Technology (NIST). Security and Privacy Controls for Information Systems and Organizations (SP 800-53 Rev. 5) : report. – Gaithersburg : NIST, 2020. – URL: <https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final>

METHOD FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING

Q.C. GUO, M.I. ZORKO, E.G. MAKEICHIK, V.Yu. TSVIATKOU

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received March 31, 2026*

Abstract. In unknown environments, a single monocular camera struggles to achieve absolute distance measurement due to the lack of scale information. This paper proposes a vision-based ranging system utilizing two mobile agents that collaborate through mutual calibration, coordinated observation, and triangulation to estimate the absolute distance to an arbitrary target. The system operates in three stages: cooperative initialization and baseline calibration, angular disparity measurement, and triangulation-based distance calculation. We specifically introduce a chain tracking measurement method to extend the effective field of view, addressing the challenge when the target and reference point are not simultaneously observable. Experimental results demonstrate that the proposed method achieves acceptable ranging errors while maintaining a favorable balance between accuracy and applicability.

Keywords: Monocular vision, collaborative ranging, triangulation, dual-agent collaboration, multi-agent system, unknown scale.

Introduction

With the widespread deployment of mobile robots and unmanned aerial vehicles in applications such as inspection, search and rescue, and warehousing, environmental perception and target localization have become critical challenges [1]. Monocular cameras are widely adopted due to their small size, low cost, and rich information output [2]; however, their inherent scale ambiguity limits their ability to perform absolute distance measurement.

Existing approaches often rely on prior maps, known target dimensions, or multi-sensor fusion (e.g., LiDAR, IMU), which face limitations in resource-constrained or unknown environments. Although binocular stereo vision can recover scale, it requires precise calibration and a fixed baseline, offering limited flexibility [3].

This paper presents a collaborative visual ranging method based on two mobile agents, each equipped with a monocular camera. The agents determine the baseline through mutual calibration [4] and then apply the law of sines using pixel positions of the target and reference points to compute the target distance. We also design a chain tracking measurement strategy [5] to overcome the limitation of the physical field of view, making the method suitable for wide-baseline scenarios. The approach is presented using a dual-agent configuration as the fundamental unit but can be naturally extended to multi-agent systems: any two agents with communication and collaboration capabilities can apply the same strategy for target ranging.

Problem Description and Model Formulation

This work aims to address the challenge that a single monocular camera cannot determine distance when the target's scale is unknown in an environment without prior information. We propose a collaborative visual ranging method using two mobile agents. The system consists of two mobile agents (Vehicle A and Vehicle B), each equipped with a monocular camera. Through mutual calibration, coordinated observation, and triangulation, the system estimates the absolute distance to an arbitrary target C.

The fundamental assumptions are as follows:

1. Both cameras A and B have been intrinsically calibrated. The camera equivalent focal length f , principal point (u_0, v_0) distortion coefficients, and horizontal field of view FOV_h are known.
2. Each camera is mounted on a square baseplate with a known side length $L \times L$. The vertical offset between the camera's optical center and the baseplate center is negligible or has been compensated for, meaning the baseplate center is considered coincident with the camera's optical center.
3. Stable communication exists between the two agents, allowing the exchange of control commands and image feature data.
4. The environment is static or quasi-static; the relative positions of A, B, and C remain unchanged during a single distance measurement process.

Methodology

The workflow of the proposed collaborative ranging system is divided into three stages: Collaborative Initialization and Baseline Calibration, Viewing Angle Difference Measurement, and Triangulation-based Distance Calculation.

1. Stage one: collaborative initialization and baseline distance calibration.

When agent A detects the target object C and sends an assistance request to B, both agents rotate in place via a control strategy until they capture each other's square baseplates in their images.

1.1. Establishing mutual gaze: Image processing algorithms (such as edge detection, Hough transform, or deep learning-based object detection) identify the imaging contour of the partner's baseplate. Control algorithms adjust the pan-tilt unit or vehicle body attitude until the image center of the partner's baseplate aligns with the agent's own image center. At this point, the optical axes of the two cameras are considered approximately coincident along the same central axis, establishing a "mutual gaze" state.

1.2. Baseline distance calculation: In the mutual gaze state, agent A (or B) calculates the imaging size of the partner's baseplate (agent B or A) through its image. Let the real side length of the square baseplate be L , and the detected pixel width of the border in the image be w (pixels). As shown in Figure 1, according to the camera's pinhole model and known intrinsic parameters, the distance from agent B's baseplate to camera A, i.e., the baseline distance AB , can be calculated as

$$AB = \frac{f \cdot L}{w}, \quad (1)$$

where f is the equivalent focal length after camera calibration.

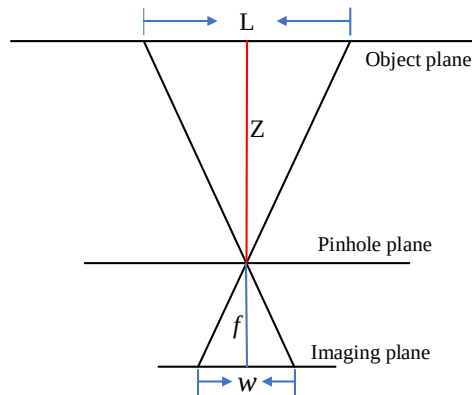


Figure 1. Pinhole Camera Model

2. Stage two: Viewing angle difference measurement.

This stage aims to accurately measure $\angle BAC$ and $\angle ABC$. The two agents take turns acting as the observer, using symmetrical methods. The measurement of $\angle ABC$ is detailed below using agent B observing A and C as an example.

- 2.1. Pixel-to-angle conversion model:

Given the camera's horizontal resolution W (pixels) and horizontal field of view FOV_h , the horizontal angular resolution per pixel θ_p is

$$\theta_p = \frac{FOV_h}{W}, \quad (2)$$

this model is the foundation for subsequent angle calculations.

2.2. Viewing angle difference measurement strategy:

Agent B starts rotating in place (e.g., clockwise) from its mutual gaze posture with A to search for target C. Depending on whether C and A can be simultaneously within B's field of view, two cases are considered:

Case 1: Co-visibility measurement.

If during rotation, both A and C appear simultaneously in B's view (as shown in Figure 2a), directly extract the horizontal pixel coordinates u_A and u_C of the centroids of A and C in the image. The viewing angle difference $\angle ABC$ is

$$\angle ABC = |u_C - u_A| \cdot \theta_p. \quad (3)$$

Case 2: Feature point chain tracking measurement.

Typically, because the baseline AB is relatively short compared to distance BC, the angular separation between points A and C in agent B's view may be very large, exceeding the camera's FOV, preventing simultaneous observation. To address this, we propose a chain tracking measurement scheme (as shown in Figure 2b). The specific process is as follows:

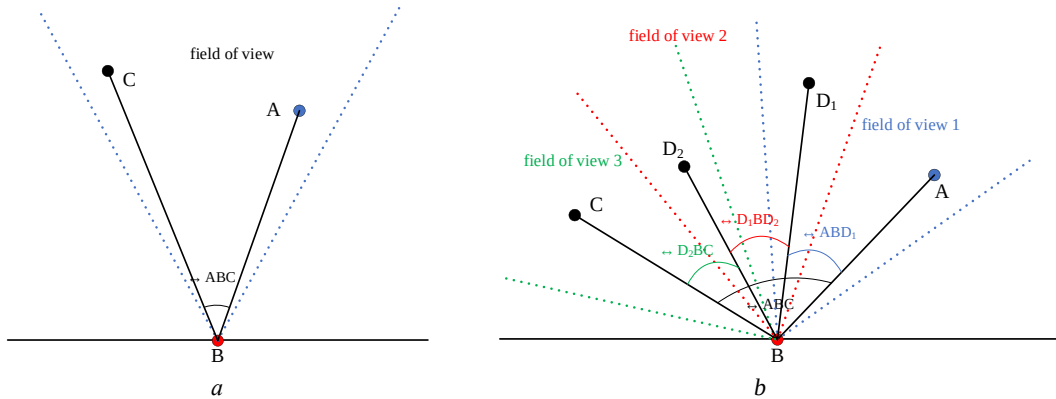


Figure 2. Visible situation: a – Simultaneously visible;
b – Feature Point Chain Tracking Measurement

(1) Agent B begins rotating. When A moves to the left edge of its view (a preset threshold) and C has not yet appeared, it searches for a stable, easily trackable natural feature point D_1 near the right edge of the view.

(2) Record the horizontal pixel coordinates u_A and u_{D_1} of A and D_1 at this moment, calculate and store the angle $\angle ABD_1$ is

$$\angle ABD_1 = |u_{D_1} - u_A| \cdot \theta_p. \quad (4)$$

(3) Agent B continues rotating, using D_1 as the new reference. When D_1 moves to the left edge of the view and C still hasn't appeared, similarly find the next strong feature point D_2 near the right edge.

(4) Record the horizontal pixel coordinates u_{D_1} and u_{D_2} of D_1 and D_2 at this moment, calculate and store $\angle D_1BD_2$ is

$$\angle D_1BD_2 = |u_{D_1} - u_{D_2}| \cdot \theta_p. \quad (5)$$

(5) Repeat this process until the target C appears within the view. Record the pixel coordinates of the last feature point D_i and C, calculate $\angle D_iBC$ is

$$\angle D_iBC = |u_C - u_{D_i}| \cdot \theta_p. \quad (6)$$

(6) Finally, $\angle ABC$ is the algebraic sum of all intermediate angles

$$\angle ABC = \angle ABD_1 + \angle D_1BD_2 + \dots + \angle D_iBC. \quad (7)$$

This method effectively extends the virtual field of view by introducing a series of intermediate natural feature points as "angular beacons," overcoming the limitation of the physical field of view.

Similarly, agent A employs the same strategy to rotate and measure $\angle BAC$.

3. Stage three: triangulation-based distance calculation

After obtaining the baseline distance AB and the two angles $\angle BAC$ and $\angle ABC$, all parameters of triangle $\triangle ABC$ are determined. According to the trigonometric sine law, the lengths of the other two sides, i.e., the distances from target C to the two agents, can be calculated as

$$AC = \frac{AB \cdot \sin(\angle ABC)}{\sin(180^\circ - \angle BAC - \angle ABC)} = \frac{AB \cdot \sin(\angle ABC)}{\sin(\angle BAC + \angle ABC)}. \quad (8)$$

Thus, we accomplish absolute distance measurement to the target object using only two collaborative monocular cameras, without prior knowledge of the object's size.

Experiments and Analysis

To validate the effectiveness of the proposed collaborative ranging method, we analyze the theoretical error of the target distance AC based on first-order error propagation theory. The comparative experiments are conducted using the baseline system parameters listed in Table 1.

Table 1. System basic parameters

Parameter (Symbol)	Value (Unit)
Agent size (L)	0.05(m)
Camera equivalent focal length (f)	3063 (pixels)
Image horizontal resolution (W)	4032 (pixels)
Camera horizontal field of view (FOV_h)	70 (deg)

1. Theoretical error model.

The ranging error of the proposed method originates mainly from baseline calibration and angular measurement, which propagate through the triangulation calculation. According to first-order error propagation theory, the maximum relative error of AC is given by

$$\delta AC_{\max} = \frac{AB}{f \cdot L} + |\cot(\angle BAC + \angle ABC)| \cdot k \theta_p + \frac{\sin(\angle BAC)}{\sin(\angle ABC) \sin(\angle BAC + \angle ABC)} \cdot k' \theta_p, \quad (9)$$

where k and k' are the chain tracking steps, and θ_p is the angular resolution per pixel (in degrees/pixel). This expression captures the combined effects of baseline length, label size, camera parameters, and geometric configuration on ranging accuracy.

For a calibrated standard binocular stereo vision system with parallel optical axes, baseline length B, and equivalent focal length f, we define a coordinate system with the left camera's optical center A as the origin: the X-axis is along the baseline direction (positive to the right), the Z-axis is along the optical axis (positive forward), and the Y-axis is perpendicular to the plane (downward). For simplicity, the target is assumed to lie in the XZ plane ($Y=0$). The coordinates of point C are (X, Z). Using first-order error propagation, the maximum relative error of the slant distance AC is

$$\delta AC_{\max_stereo} = \frac{|X| \cdot Z}{2f(X^2 + Z^2)} + \frac{Z}{f \cdot B}. \quad (1)$$

2. Experimental comparison and analysis.

To assess the performance of the proposed method (dual-agent collaborative monocular vision) and binocular stereo vision in unknown environments, we conducted large-scale Monte Carlo simulations. The experiment comprised 100,00 independent trials, each generating 20 random triangles satisfying geometric constraints (baseline 50 - 500 cm, target distance 100 - 1000 cm, angles 10° - 170°), resulting in a total of 200,000 scenarios. The measurability of the binocular system and the theoretical relative errors of both methods were statistically analyzed.

Regarding applicability, the proposed method demonstrates a significant advantage over traditional binocular stereo vision. As illustrated in Fig. 3, there are two types of scenarios where the binocular system fails because the target lies outside the common field of view.

Statistical results show that the average proportion of measurable scenes for the binocular system is only 16.42%, with a standard deviation of 8.36%. In the worst case, none of the 20 scenes in a trial were measurable (0% measurability), while the best case achieved only 55.0% (Fig. 4a). This indicates that in typical unknown environments, the applicability of binocular stereo vision is severely limited by the common field-of-view constraint, with over 80% of random geometric configurations being unsuitable for binocular ranging. In contrast, the proposed method achieves 100% scene measurability through its chain tracking strategy.

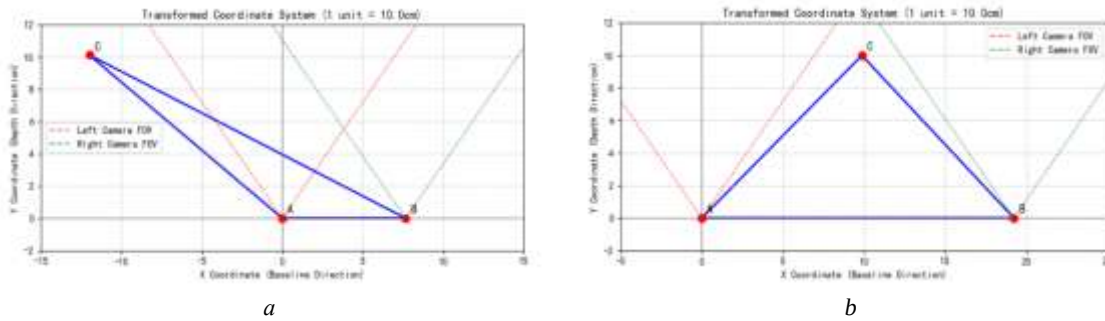


Figure 3. Situations that cannot be measured under a binocular system

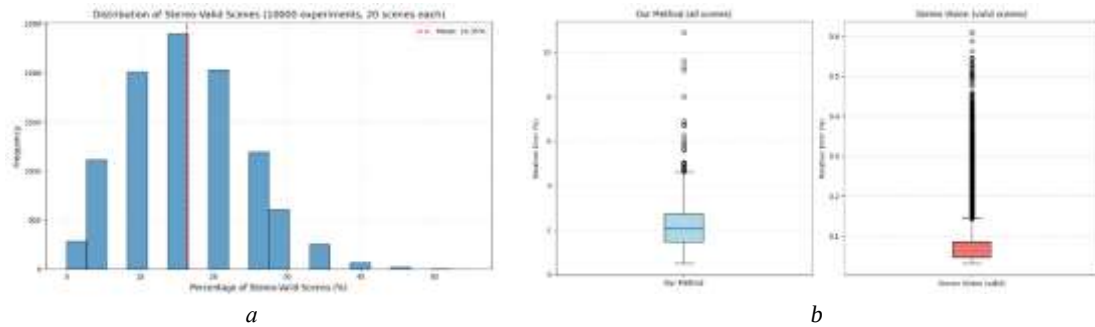


Figure 4. Experimental results: a – the proportion that can be measured by the binocular system; b – error rate

In terms of accuracy, the proposed method successfully estimated distances in all 200,000 scenarios without failure. Figure 4b shows that the error distribution of the proposed method is centered around 2%, with the median close to the mean, indicating a symmetric distribution. The error range spans 0.48% to 10.89%, covering typical scenarios from close range to long distance and various angles. This accuracy meets the basic requirements of many robotic tasks (e.g., obstacle avoidance, target approach). For scenes where binocular vision is applicable, its error is extremely small (mean 0.076%) with a highly concentrated distribution (median 0.062%), demonstrating high and stable accuracy under favorable geometric conditions. However, its main limitation remains the restricted set of measurable scenes, while the proposed method achieves a favorable balance between accuracy and applicability.

Conclusion

This paper presents a monocular vision ranging system based on dual-agent collaboration, capable of absolute distance estimation for arbitrary targets in unknown environments. The system operates through three stages: cooperative initialization and baseline calibration, angular disparity measurement, and triangulation-based distance calculation, enabling ranging without prior knowledge of the target's size. Notably, the proposed chain tracking measurement strategy effectively extends the virtual field of view, addressing the challenge when the target and reference point cannot be observed simultaneously, thus making the method suitable for wide-baseline applications.

The method is described using a dual-agent configuration as the minimal implementation unit, but its core mechanisms (mutual calibration, collaborative observation, and triangulation) are fully applicable to multi-agent systems. In such systems, any two agents with communication and collaboration capabilities can apply the same strategy for collaborative ranging, thereby forming a distributed collaborative perception network and enhancing perception coverage in large-scale unknown environments.

Future work will focus on real-time collaborative ranging in dynamic scenes, multi-agent collaborative optimization, and the deployment and validation of the method on practical robotic platforms.

References

1. Lyu M., Zhao Y., Huang C., et al. Unmanned aerial vehicles for search and rescue: A survey. // Remote Sensing. 2023. Vol. 15. P. 3266.
2. Guo Q. Implementation principle of monocular ranging. // Технологии передачи и обработки информации : материалы Международного научно-технического семинара. Минск, 2024. P. 78–82.
3. Guo Q. Research on relative localization of mobile robots based on monocular vision and known tags. // Технологии передачи и обработки информации : материалы Международного научно-технического семинара. Минск, 2025. P. 141–143.
4. Guo Q. Research on depth estimation for monocular cameras based on prior information. // Технологии передачи и обработки информации : материалы Международного научно-технического семинара. Минск, 2025. P. 179–182.
5. Guo Q. Method for extended field-of-view angle measurement based on chain tracking of feature points. // XII Белорусско-Китайский молодежный инновационный форум «Новые горизонты - 2025». Минск, 2025. P. 72–74.

SENSOR FUSION FOR AUTONOMOUS VEHICLE ENVIRONMENTAL PERCEPTION: A REPRODUCIBLE COMPARISON

XINYU ZHANG

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 10, 2026

Abstract. Autonomous vehicles (AVs) rely heavily on advanced sensor technologies to ensure safe navigation and obstacle avoidance in dynamic and diverse environments, yet single sensors have inherent limitations in perceptual information completeness, environmental adaptability and functional safety, failing to support stable autonomous driving in complex open scenarios. Sensor fusion thus becomes a core approach for autonomous driving perception systems, which synthesizes multi-source sensory data to boost perception accuracy, enrich environmental cognition and enhance system robustness beyond individual sensor capabilities. This paper reproduces and compares two different sensor fusion methods, intending to offer practical references for the real-world application of such systems in autonomous driving.

Keywords: autonomous vehicles, sensor fusion, object detection.

Introduction

As a core component of intelligent transport systems, autonomous driving technology hinges on precise, comprehensive and robust perception of the driving environment, with the reliability of environmental perception directly determining the decision-making safety and operational stability of autonomous driving systems [1]. Single-sensor perception solutions have inherent limitations rooted in their physical characteristics: RGB cameras capture rich visual semantic information for accurate identification of object categories, textures and scenes, yet they are vulnerable to environmental interferences such as lighting changes, object occlusion and adverse weather, and cannot directly obtain the 3D geometric and distance information of objects [2]. In contrast, LiDAR generates high-precision 3D point cloud data via laser ranging to realize illumination-independent detection of target distance and spatial position with strong anti-interference ability [3], but it lacks target semantic information, making it hard to distinguish the category attributes of detected objects [4], and its point cloud data is plagued by high noise and sparse features for small targets.

These inherent drawbacks render single-sensor solutions incapable of meeting the perception demands of autonomous driving in complex scenarios, thus making multimodal fusion perception a key technological approach to address this issue [3]. Fusing the visual semantic information of RGB cameras with the 3D geometric information of LiDAR leverages the complementary strengths of the two sensors, compensates for the information gaps of monomodal perception [2], enhances the completeness, accuracy and robustness of environmental perception, and provides more reliable environmental data support for the decision-making and planning modules of autonomous driving systems, which has become a research focus and development trend in the field of autonomous driving perception [5].

In recent years, extensive research has been conducted on vision-LiDAR multimodal fusion perception, with fusion levels evolving from raw data-level to feature-level and decision-level, and fusion algorithms advancing from traditional manual feature fusion to deep learning-based end-to-end fusion [6]. Although existing studies have verified the effectiveness of multimodal fusion perception, most focus on the algorithm optimization of a single fusion scheme, with few comparative analyses of different-level fusion schemes and a lack of intuitive validation of their detection performance in real-world scenarios. Additionally, experimental research on multimodal fusion for small and medium-sized

datasets is relatively scarce, which fails to meet the research needs of university academic research and introductory engineering practice.

To fill these gaps, this paper takes the KITTI [7] autonomous driving dataset as the experimental carrier and constructs a vision-LiDAR multimodal fusion perception experimental framework to conduct end-to-end experimental verification from monomodal perception to multimodal fusion perception. The research mainly includes four aspects: first, building RGB-only and LiDAR-only monomodal perception modules to verify their inherent limitations in acquiring semantic and geometric information; second, designing and implementing a data-level (low-level) fusion scheme based on point cloud projection, which projects LiDAR point clouds onto the image plane to achieve pixel-level fusion of visual pixel and LiDAR depth information for raw data-level information complementarity; third, designing and implementing a feature-level (middle-level) fusion scheme based on feature extraction and weighted fusion [4], which adopts lightweight encoders to extract image semantic features and point cloud geometric features respectively, and realizes high-level feature alignment and fusion for feature-level information interaction; fourth, systematically comparing and analyzing the experimental results of monomodal perception and different-level multimodal fusion perception from multiple dimensions including the number of detected targets, category recognition ability, distance information acquisition accuracy and perception robustness. Through this research, the technical advantages and application limitations of each perception scheme are explored in depth, providing experimental evidence and technical references for the engineering design and introductory research of vision-LiDAR multimodal fusion perception algorithms.

Related Works

Monomodal perception methods are divided into visual perception and LiDAR perception. The former relies on camera-based object detection (such as YOLO and Faster R-CNN), which excels at semantic recognition but lacks depth information and is susceptible to lighting conditions or occlusion; the latter (DBSCAN) excels at geometric localisation and ranging but lacks semantic information [8].

Multi-sensor fusion can be categorised into three types based on the level of fusion. Firstly, there is data-level fusion, which occurs at the raw data level (such as projecting point clouds onto images); whilst this preserves the most complete information, it involves significant computational effort and demands high registration accuracy. Secondly, there is feature-level fusion, which operates at the feature level (matching 2D detection boxes with 3D clusters). This approach balances accuracy and efficiency and is the current mainstream solution. Finally, there is decision-level fusion (high-level fusion), which operates at the decision-making level (voting on results from different modalities). This approach offers high flexibility but results in significant information loss.

Methodology

This research implements four perception modes: object detection using RGB cameras only, clustering of LiDAR point clouds only, data-level fusion and feature-level fusion. The overall processing workflow of the proposed multi-sensor fusion perception method is as follows: it sequentially performs the loading of multi-source sensor data; camera and LiDAR external parameter calibration and coordinate registration; single-modal perception comprising RGB image object detection and LiDAR point cloud clustering; and multi-modal information fusion at the raw data and intermediate feature levels, ultimately resulting in the visualisation of the fusion results.

1. Data and Sensor Configuration

The experiment uses the KITTI dataset, which includes: colour camera images, LiDAR point clouds, calibration files. Data formats: Images, PNG format, 2D object detection input. Point clouds, BIN format, with each point containing (x, y, z, r) , where r is the reflectance. Calibration parameters, external parameters matrix R , translation vector T .

2. Sensor Calibration and Coordinate Transformation

2.1. Conversion from lidar to camera coordinate system

Converting LiDAR points from the LiDAR coordinate system to the camera coordinate system, the calculation formula is shown as equation

$$p_{cam} = R \cdot P_{velo} + T. \quad (1)$$

2.2. Projection from the camera onto the image plane

Use the projection matrix P_{rect_03} from the calibration file to project 3D points onto 2D pixel coordinates.

3. Unimodal Perception Module

3.1. RGB camera object detection

2D object detection using the YOLOv8 lightweight network [9], input RGB image and output object class, confidence score and 2D bounding. The purpose of this step is to provide semantic information about objects (vehicles, pedestrians, road signs, etc.).

3.2. LiDAR point cloud clustering

Extracting 3D targets using DBSCAN unsupervised clustering. The input is a ground-corrected LiDAR point cloud, and the output is a number of point cloud clusters, each representing a distinct target. The purpose of this step is to provide the 3D positions and distances of the targets. In experiments using LiDAR alone, the system can output precise distances but is unable to identify target categories.

4. Low-Level Fusion

The low-level fusion performs alignment directly at the raw data level. First, the image and point cloud are read; the point cloud is then projected onto the image via coordinate transformation and projection. This is subsequently fed into the YOLO detector to obtain 2D bounding boxes; the mean depth of all projected points within the box is calculated; and finally, the object class and average depth are output. The depth is calculated using the formula as equation

$$d_{mean} = \frac{1}{N} \sum_{i=1}^N d_i, \quad (2)$$

where N represents the number of valid points in the detection box. This fusion method preserves the original depth distribution without point cloud denoising or clustering; it is simple to implement and retains all information, but is susceptible to interference from background points, and the depth accuracy is generally moderate.

5. Middle-Level Fusion

The intermediate fusion performs matching at the feature level: the RGB component is responsible for YOLO's output of 2D bounding boxes and classes, whilst the LiDAR component is responsible for DBSCAN's output of 3D cluster centres. Subsequently, the 3D cluster centres are projected onto the image for matching; if a bounding box contains a cluster centre, it is identified as the same object. Finally, the object class and the precise depth of the centre are output. This approach combines cluster-based noise filtering with ground control points, utilising a depth map centred on the target, thereby achieving greater accuracy whilst offering both semantic and geometric advantages.

Ablation Study

To demonstrate the limitations of a single sensor and the benefits of multi-sensor fusion, this paper conducted ablation experiments on the KITTI dataset, designing four comparison scenarios: RGB vision only, LiDAR only, fusion at the raw data level, and fusion at the feature level.

1. RGB-only

When using only RGB images for YOLOv8 object detection, the model is able to identify objects such as cars and parking meters within the scene. However, due to the lack of depth information, the model is unable to provide the three-dimensional position or distance parameters of the objects. At the same time, visual inspection is susceptible to changes in lighting, occlusion and shadows, resulting in insufficient stability in object recognition. This experiment demonstrated the fundamental limitations of RGB-only: whilst it possesses semantic capabilities, it lacks the ability to determine distance and geometric positioning.

2. LiDAR-only

Using only LiDAR point clouds for DBSCAN clustering, the experiment partitioned the raw point cloud into 354 target clusters. A bird's-eye view (BEV) visualization (Figure 1) shows that the point

cloud exhibits a typical distribution of scanning arcs, with each cluster distinguished by a different colour.

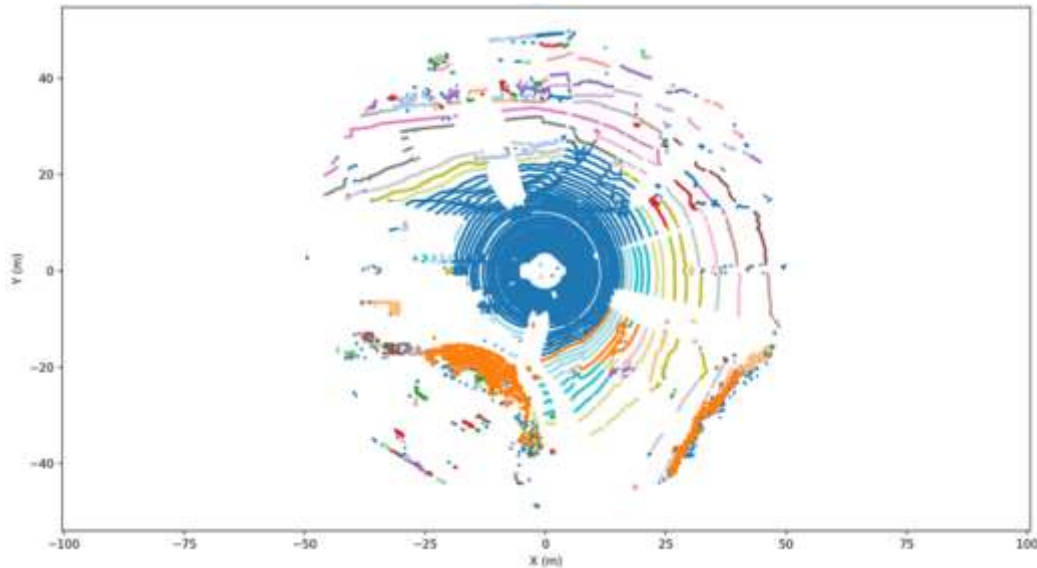


Figure 1. LiDAR obstacle clustering results

This approach provides precise 3D position and distance information for targets, but lacks semantic classification and cannot distinguish whether a target is a vehicle, a pedestrian or a road sign. Consequently, a pure LiDAR approach can only provide a geometric foundation and cannot achieve target-level semantic understanding. Experimental results indicate that a single LiDAR modality suffers from a significant lack of semantic information in target recognition.

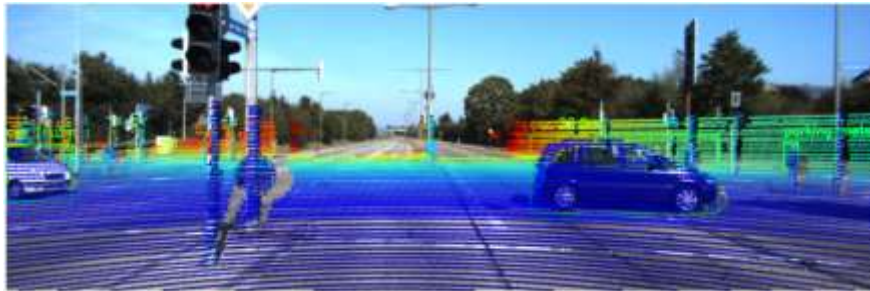


Figure 2. In-depth visualisation of data-level Fusion

3. Data-Level Fusion (low-level fusion)

The low-level fusion projects the LiDAR point cloud directly onto the image plane, thereby achieving spatial alignment of the raw sensor data. This model is capable of outputting target categories and approximate depths: the depth of the vehicle and that of the parking timer (Figure 2). Although this method preserves the full scene depth, it does not perform clustering on the point cloud; consequently, a large number of background points are included within the detection boxes, leading to noise interference in the depth calculations.

4. Feature-Level Fusion (middle-level fusion)

The middle-level fusion stage first extracts visual features (YOLOv8 2D detection) and point cloud features (DBSCAN 3D clustering) separately, and then achieves multimodal information fusion through feature matching (Figure 3).

Compared with other approaches, feature-level fusion offers three key advantages: Firstly, more precise target localisation: by filtering out noise through clustering, only the target's central point cloud is retained, resulting in distances that are closer to the true values; Secondly, greater robustness: only targets successfully matched across modalities are retained, filtering out isolated false positives; Thirdly, complementary semantic and geometric information: it combines visual class recognition with the precise ranging capabilities of LiDAR.



Figure 3. Feature-level Fusion obstacle detection

Result and discussion

1. Results of monomodal perception

When relying solely on an RGB camera for object detection, the model identified a total of six objects, comprising two cars, three traffic lights and one parking meter. Visual detection is capable of clearly distinguishing between object categories, but it cannot provide information on the distance to the objects and is susceptible to environmental factors such as lighting and occlusion; consequently, the confidence level for detecting small or distant objects is relatively low. However, when relying solely on LiDAR for point cloud clustering, a total of 354 3D target clusters were obtained. Whilst LiDAR provides precise distance information and offers more stable target detection at long ranges and in low-light conditions, it is unable to distinguish between target categories. Consequently, the clustering results contain a large number of noise points and non-target clusters, resulting in a significant lack of semantic information regarding the targets.

2. Results of multimodal fusion perception

In the initial fusion stage, the LiDAR point cloud is projected onto the image plane to achieve pixel-level fusion of depth information with visual information. This method yields a total of 184999 valid projected points, enabling the assignment of precise distance information to visually detected targets: The average distances to the two vehicles are 22,01 m and 18,29 m respectively, the average distance to the parking timer was 28,89 m.

It simultaneously retains the class information from visual detection and the distance information from LiDAR, achieving a preliminary alignment of semantic and geometric information for the targets.

After performing weighted fusion of image features and point cloud features at the feature level, the feature-level fusion results are fed into YOLOv8 for object detection. Ultimately, a total of two car objects were identified; the detection results were consistent with those of single-modal visual detection, whilst also achieving more reliable object distance estimates. Target 1 (green box): car, 0,92 confidence, 13,13m, labelled as Vision and LiDAR fusion type. Target 2 (orange box): car, 0,63 confidence, unknown distance, labelled as Vision-only type.

The fused model significantly improves the accuracy of distance estimation for close-range targets whilst retaining the high-precision class recognition capability of visual detection, thereby achieving complementary information compared to single-modal perception (as shown in Table1).

Table 1. Performance Comparison of Different Perception Methods

Methods	Target	Ctegrory recognition	Distance information	Advantages	Limitation
RGB-only	6	Yes	No	clear category recognition	no distance information
LiDAR-only	354	No	Yes	accurate distance measurement	no category information
Low-level Fusion	3	Yes	Yes	pixel-level information alignment	relies on precise calibration
Mid-level Fusion	2	Yes	Yes	feature-level information complementarity and stronger robustness	complex implementation

In the current experiment, middle-level fusion is still unable to provide distance information for some targets (such as Target 2). The primary reason for this is the absence of a valid LiDAR point cloud projection for that target area, which prevents the acquisition of geometric information during feature fusion. In future, the completeness and accuracy of the fused perception can be further improved by optimising the point cloud feature generation strategy and refining the feature alignment algorithm.

Conclusion

The experimental results demonstrate that: Single-modal perception has clear limitations: visual detection lacks geometric information, whilst LiDAR lacks semantic information; neither can meet the requirements of autonomous driving for comprehensive target perception. Low-level fusion achieves basic information complementarity: assigning distance information to visual targets via point cloud projection is the most intuitive fusion method, but it relies on precise sensor calibration and incurs high computational costs. Mid-level fusion enhances perception robustness: by fusing the strengths of the two modalities at the feature level, this approach retains the class recognition capabilities of visual detection whilst incorporating the distance information from LiDAR, significantly improving the perception accuracy of close-range targets. It is a fusion scheme of greater engineering value.

References

1. Y. Zhang, S. Lu.// SEC 2025.
2. M. Nawaz, J. K. T. Tang, K. Bibi, S. Xiao, H. P. Ho, W. Yuan.// IEEE 2024.
3. D. J. Yeong, G. Velasco-hernandez, J. Barry, J. Walsh.// Sensors 2021.
4. F. García, D. Martín, A. de La Escalera, J. M. Armingol.// IEEE 2017.
5. W. Li, Z. Xie, P. He.// ICSI 2025.
6. B. S. Jahromi, T. Tulabandhula, S. Cetin.// Sensors (Switzerland). 2019.
7. IEEE Staff.// 2012 IEEE Conference on Computer Vision and Pattern Recognition 2012.
8. W. Hadikurniawati, K. D. Hartomo, I. Sembiring.// ICMERALDA 2023.
9. I. Logeswaran, H. Eissa, R. Almadhoun.// ICAC 2024.

УДК 621.391

СОВЕРШЕНСТВОВАНИЕ ОРГАНИЗАЦИОННЫХ МЕХАНИЗМОВ УПРАВЛЕНИЯ РАЗРАБОТКОЙ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ В ИНФОРМАЦИОННО-РАСЧЁТНОМ ЦЕНТРЕ (ИРЦ) РУП «БЕЛТЕЛЕКОМ» НА ОСНОВЕ ТЕОРИИ АКТИВНЫХ СИСТЕМ

Д.М. ГОРБУКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 17 апреля 2026

Аннотация. В работе рассмотрены теоретические и практические аспекты управления разработкой программного обеспечения в информационно-расчётном центре (ИРЦ) РУП «Белтелеком» на основе теории управления организационными системами и концепции активности агентов. Проведён анализ существующих механизмов управления (планирование, мотивация, контроль), выявлены их недостатки, связанные с возможным искажением информации и отклонением от планов. Предложены и обоснованы мероприятия по совершенствованию организационных механизмов: автоматизация управления на базе Jіga, централизация функций учёта и контроля, внедрение конкурентного стимулирования программистов. Показано, что реализация предложенных мер позволяет повысить результативность системы управления производством ИРЦ. Также рассмотрены возможности применения синергетического подхода и самоорганизации для дальнейшего развития механизмов управления.

Ключевые слова: управленческий цикл, управление организационными системами, активные системы, механизмы управления, разработка программного обеспечения, ИРЦ РУП «Белтелеком», Jіga, мотивация, контроль, синергетика, самоорганизация.

Введение

Актуальность. Цифровая трансформация РУП «Белтелеком» требует создания команды квалифицированных инженеров-программистов и эффективного управления ею для разработки и сопровождения информационных систем силами информационно-расчётного центра (ИРЦ). Однако сотрудники (активные агенты) обладают собственными интересами, могут исказить информацию или отклоняться от планов, что снижает результативность управления. Классические административные методы здесь недостаточно эффективны.

Проблема. В системе управления производством ИРЦ отсутствуют механизмы, в полной мере учитывающие активность сотрудников, что приводит к задержкам, снижению качества ПО и росту управленческих издержек.

Объект исследования – система управления производством ИРЦ РУП «Белтелеком».

Предмет исследования – организационные механизмы управления производством ИРЦ (планирование, мотивация, контроль).

Цель работы – разработка рекомендаций по повышению результативности системы управления разработкой ПО в ИРЦ на основе механизмов, учитывающих активность сотрудников.

Задачи:

1. Изучить теоретические основы управления активными системами и механизмами управления.
2. Проанализировать действующую систему управления ИРЦ и выявить её недостатки.
3. Предложить и обосновать усовершенствования механизмов управления (автоматизация, централизация контроля, конкурентное стимулирование).

4. Оценить эффективность предложенных мероприятий.

Методология. Теория управления организационными системами (Бурков, Новиков и др.), концепция активности агентов, управленческий цикл СПРУКАР.

Практическая значимость – предложенные механизмы могут быть внедрены в ИРЦ и аналогичных подразделениях, обеспечивая рост прозрачности, мотивации и дисциплины.

Теоретические аспекты исследования организационных механизмов управления

При описании механизмов управления удобно использовать типовой шаблон (в каждом случае выбираются конкретные, соответствующие рассматриваемому механизму характеристики и пиктограммы) [3, с.101-103]. Данный шаблон представлен на рисунке 1.

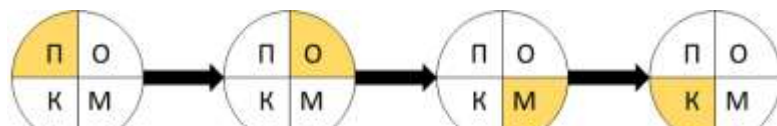
Название механизма		
Краткое описание; библиографические ссылки на подробное описание.		
Структура организационной системы	Рисунок в соответствии с шаблоном – используется одна из схем, приведенных на рис. 3-5 с содержательной детализацией.	Расширения базовой организационной системы
Предмет управления		Функция управления
Эффект от внедрения механизма		
Состав организационной системы		
Классификация типовой структуры организационной системы	 – базовая (один Центр, один агент);  – всеерная (управление подразделением);  – ромбовидная структура (распределенный контроль);  – многоуровневая иерархия;  – матричная структура управления	
Действия, ограничения и нормы деятельности участников		
Цели и предпочтения участников	 – схематичные представления функций «прибыли» Центра и агентов с их содержательной детализацией	
Информированность участников	 – какой информацией обладают участники ОС на момент принятия ими решения.	
Предмет управления	– Состав организационной системы (управление составом) – Структура организационной системы (управление структурой) – Ограничения и нормы деятельности (институциональное управление) – Предпочтения (мотивационное управление) – Информированность (информационное управление)	
Расширения базовой организационной системы	– Число периодов функционирования (динамические ОС) – Число агентов и Центров (многоэлементные ОС) – Структура (многоуровневые ОС, ОС с распределенным контролем) – Информированность (ОС с неопределенностью, ОС с сообщением информации)	
Функция управления	Планирование → Организация → Мотивация → Контроль (стимулирование) 	

Рис. 1 – Типовой шаблон описания механизма управления
Примечание – Источник: собственная разработка на основе [3, с. 101-103]

Задача управления	– Мониторинг и анализ текущего состояния – Прогноз развития ситуации – Целеполагание – Выбор технологии деятельности – Планирование и распределение ресурсов – Стимулирование (мотивация) – Контроль и оперативное управление – Анализ и улучшения деятельности, рефлексия
Область применения	
Условия применимости	
Применение совместно с другими механизмами	
Необходимая информация и алгоритм применения механизма	
Вход - выходная схема	
Дополнительная информация	Этот раздел шаблона содержит подробное описание механизма управления, примеры расчетов его параметров, исторические справки, ссылки на дополнительную литературу и прочую полезную информацию.

Рис. 1, лист 2

Для описания функционирования систем управления используется понятие управленческого цикла, то есть модели, описывающей процесс управления как циклическое повторение типовых этапов. На сегодняшний день известно несколько десятков моделей описания управленческого цикла, в том числе:

- цикл А. Файоля [3, с. 38];
- цикл У. Деминга PDCA (Plan – планируй, Do – делай, Check – проверяй, Act – воздействуй) [2, с. 14];
- комплекс задач и функций управления [3, с. 26];
- цикл СПРУКАР [3, с. 38].

Этапы (функции) управленческого цикла СПРУКАР:

1. Сбор и анализ информации для выработки и принятия управленческого решения.
2. Планирование действий – разработка и принятие управленческого решения.
3. Реализация – организация исполнения (в том числе, мотивация исполнителей), и исполнение управленческого решения.
4. Учет результатов реализации управленческого решения.
5. Контроль – сравнение факта с намеченными результатами.
6. Анализ – выявление и изучение причин отклонения от намеченных результатов.
7. Регулирование хода исполнения управленческого решения, коррекция (при необходимости) ранее принятых решений, применение к исполнителям предусмотренных мотивационных воздействий с целью усиления выявленных положительных тенденций и нейтрализации отрицательных (рисунок 2) [3, с. 39].



Рис. 2 – Управленческий цикл СПРУКАР

Примечание – Источник: собственная разработка на основе [3, с. 40]

Порядок функционирования. Можно детализировать порядок функционирования и взаимодействия субъекта и объекта управления. Описание порядка функционирования с однозначным разнесением операций между участниками называется процедурой. Если прописать связи между подфункциями функций управления, то в результате получится описание управленческих процедур. Часть связей в составе процедуры будет относиться к действиям субъекта управления, часть связей – к действиям объекта управления, а часть – к взаимодействию субъекта и объекта управления. Совокупность процедур, регламентирующих действия субъекта управления (принимаемые им решения), составляет механизм управления [3, с. 41]

Организация является достаточно сложным объектом управления. Способность объектов управления или агента к целеполаганию, самостоятельному выбору действий отражает концепция активности объектов управления (агентов) в организационных системах. В соответствии с ней агенты имеют определенные возможности и простор для самостоятельного целеполагания и принятия решений о своих действиях (в рамках полноценной структуры деятельности, с учетом всех процессуальных ее компонент). Проявлениями активности являются: искажение информации, выбор состояния, не совпадающего с планом, недобросовестное поведение и т.д. – все эти возможные эффекты должны быть учтены при разработке и применении механизмов управления. Для эффективного управления такими объектами необходимо моделировать их поведение, то есть прогнозировать их реакцию на те или иные управляющие воздействия.

Концепция максимизации полезности представляет собой модель «поведения экономического человека», далее – агента. Агент ведет себя (сообщает информацию, выбирает действия и т.д.) так, чтобы максимизировать свою полезность.

Выделяют два вида экономических агентов – субъектов экономики:

- хозяйствующий субъект рынка (примеры – организация, холдинг, фирма, корпорация);
- сотрудник(и) хозяйствующего субъекта.

Структура деятельности активного агента выглядит следующим образом (рисунок 3):

- агент описывается информированностью (той информацией, которой он обладает о существенных параметрах ОС и внешней среды), ограничениями, предпочтениями и действием;
- на его вход (стрелка слева) поступает, во-первых, информация об окружающей среде и других агентах и, во-вторых, управляющие воздействия в виде выбранного Центром механизма управления (стрелка сверху);
- «выходом» агента (его действием) является сообщаемая Центру и/или другим агентам информация (пунктирная стрелка справа) и выбранное им действие (жирная стрелка справа).

Такая нотация используется ниже в качестве элементарного блока при описании механизмов управления организационными системами (ОС) [3, с. 57-58].

Действием Центра является выбор механизма управления – зависимости управляющих воздействий от действий агента.



Рис. 3 – Структура деятельности активного агента
Примечание – Источник: [3, с. 58]

Ключевое отличие Центра от Агента заключается в том, что он обладает властью – имеет право «сделать ход первым», установив для агента условия деятельности (механизм). Полезность Центра обычно зависит от действий агента, то есть эффективность управления определяется полезностью Центра (представляющего интересы организационной системы в целом) от деятельности агента [3, с. 59].

В моделях управления организационным поведением каждый экономический агент описывается четырьмя базовыми параметрами, агрегирующими процессуальные компоненты его деятельности:

- 1) ограничения и нормы деятельности (процессуальные компоненты деятельности – условия, нормы, принципы, технология);
- 2) функция полезности (процессуальные компоненты деятельности – потребности, мотивы, цели, задачи, критерии);
- 3) информированность (процессуальные компоненты деятельности – информация о внешней среде, включающей, в том числе, других агентов, взаимодействующих с данным);
- 4) действие, выбираемое агентом на основании своей информированности с учетом ограничений и своей полезности. Действие является показателем состояния агента и в значительной степени определяет результат его деятельности.

С экономическим агентом связывают два вида возможных действий (в рамках существующих ограничений):

- сообщение субъекту управления и/или другим агентам информации о неопределенных для них параметрах;
- действие по выбору своего состояния (объема производимой продукции, затрачиваемого времени и т.д.).

Двум видам возможных действий агента соответствует цикл функционирования системы управления, содержащий два контура управления соответствующих планированию и реализации. В составе цикла управления действуют несколько классов механизмов управления организационным поведением – механизмы планирования, организации, стимулирования и контроля. Они объединяются в комплексный механизм функционирования организационной системы в рамках цикла управления [3, с. 67].

Выбор механизма управления является действием Центра (рисунок 4):

- информационное управление (воздействие на информированность);
- институциональное управление (воздействие на ограничения и нормы деятельности);

– мотивационное управление (воздействие на предпочтения).



Рис. 4 – Выбор механизма управления является действием Центра
Примечание – Источник: собственная разработка на основе [3, с. 67]

Общий (для всех этапов цикла управления) порядок функционирования организационной системы (при фиксированном механизме управления) изображен на рисунке 5:

I этап: Центр делает первый ход – сообщает агенту механизм функционирования («правила игры») в общем виде. Например, зависимость количества выделяемого ресурса от сообщенной потребности в нем или зависимость размера вознаграждения от достигнутых результатов (состояния агента).

II этап: агент сообщает Центру информацию о неизвестных последнему параметрах (например, подает заявку на ресурс или сообщает информацию о своих предпочтениях).

III этап: Центр сообщает агенту (детализирует) параметры механизма функционирования. Например, назначает план – ожидаемый Центром результат деятельности агента.

IV этап: агент выбирает действие, реализуется результат деятельности.

V этап: Центр получает информацию о действии агента

VI этап: Центр в соответствии с механизмом функционирования сообщает агенту свое действие – например, количество выделенного ресурса, размер вознаграждения и т.д.

Этапы I-III соответствуют циклу планирования, этапы III-VI – циклу реализации. Нижний прямоугольник на рисунке 5 представляет собой модель экономического агента [3, с. 61]



Рис. 5 – Типовая схема субъекта управления
Примечание – Источник: [3, с. 61]

В своей статье «Теория управления и хаоса, теория сложности и теория систем самоорганизации» группа исследователей Папа Мбенге, Джеймс Ондрачек, М. Саид, Энди Берч [17] из Государственного университета Майнот, Северная Дакота, США представляет краткий обзор теории хаоса, теории сложности и теории самоорганизующихся систем и их применения к менеджменту.

В современных организациях и фирмах интерактивное поведение индивидов формирует систему, а система, в свою очередь, непропорционально описывает поведение индивидов внутри нее. Другими словами, вся система больше, чем сумма ее частей. Необходимость определения таких систем породила теорию сложности, которая собирает и анализирует данные на разных уровнях детализации (от отдельных лиц, групп, обществ). Определяемая как набор взаимосвязанных видов деятельности, которые взаимодействуют нетривиальными способами, которые трудно предсказать, организация или фирма становится сложной системой, которая может извлечь выгоду с точки зрения теории сложности. Однако авторы данной статьи не приводят каких-либо интересных примеров

Важным разделом теории самоорганизующихся систем является синергетика. Синергетика описывает фазы переходов и формирование закономерностей в системах, где неустойчивые режимы доминируют над устойчивыми и определяют развитие динамики систем.

Авторы статьи приходят к выводу, что в современных условиях управления постепенных изменений уже недостаточно, поскольку события, угрозы и возможности не только приходят в организацию быстрее или с меньшей предсказуемостью, но также сходятся и влияют друг на друга, создавая полностью уникальные ситуации. Эти условия требуют адаптивности и новых способов руководства. С точки зрения менеджмента теории хаоса, теории сложности и теории самоорганизующихся систем находятся на стадии развития пред-парадигмы. Этого недостаточно для формирования программы теоретических исследований в области менеджмента на базе этих теорий. Однако, они дают полезные описания явлений в менеджменте, которые могут эффективно направлять усилия практиков по экспериментированию с этими теориями для поиска нового понимания управления.

В своей статье «Организация и самоорганизация» исследователи В. Чехер из Бернского университета (Швейцария) и Э. Дж. Бруннер из Института педагогических наук Тюбингенского университета, Германия [18] обсуждают системный подход к организационной теории на основе теории самоорганизации.

Организации рассматриваются как нелинейные системы, характеризующиеся сложностью на микроскопическом уровне, циклической причинностью и открытостью психосоциальному воздействию внешней среды. Сложные системы нередуцируемы и вызывают сложное поведение, которое трудно, если не невозможно, предсказать.

Из феноменологии самоорганизующихся систем авторы статьи выводят две гипотезы: система всегда больше суммы своих частей; Управленческий контроль можно разделить на управление влиянием и событийное управление.

Исследователи приходят к выводу, что, если система находится вблизи точки бифуркации, шансы менеджера повлиять на систему высоки. Менеджер в организации вполне может сам внести решающий сдвиг. Прогноз развития организации ненадежен, но на нее пока легко оказать влияние.

Авторы статьи также утверждают, что обратная ситуация имеет место в областях пространства параметров, удаленных от критических точек. Здесь прогноз прост, потому что система в организации стабильна. Именно поэтому вмешательство менеджера может встретить «сопротивление» системы.

Авторы статьи предлагают рассматривать влияние коммуникаций в организации как феномен «стрелы времени», связанной с диссипацией. Они утверждают, что в социальных и ментальных системах нарушение симметрии необратимо. Коммуникации невозможно ни вернуть, ни инвертировать, потому что система моментально меняется при любом событии.

Исследователи задаются вопросом, как менеджеру справиться с хаосом в системе. В качестве практического вывода они предлагают два подхода:

1. Управление обязательно должно быть косвенным; управление осуществляется путем обеспечения основы для самоорганизации. Руководство работает над параметрами контроля системы. Управление не может напрямую определять равновесие системы.

2. Авторы выражают уверенность в проявлении явления, для которого используется термин «синергия», выражая уверенность в том, что выгоды можно ожидать от взаимодействия ранее независимых частей. Например, исследователи предлагают добиться синергетического эффекта посредством неформальных встреч.

В своей статье «Механизмы управления организационным поведением: исследование» коллектив исследователей В.Н. Бурков, М.В. Губко, Н.А. Коргин, Д.А. Новиков [16] из Института проблем управления РАН, Москва, Россия рассматривает основы версии теории проектирования механизмов, адаптированной для решения проблем управления.

Очень интересны результаты исследования авторов статьи по вопросам управления сотрудниками организации. Они рассматривают сотрудников организации как элементы активных систем. Элементами активных систем являются люди, имеющие индивидуальные интересы, способные самостоятельно выбирать действия и манипулировать информацией. Современные тенденции в менеджменте рассматривают рабочую силу, обладающую поведением, способным к саморегуляции и саморазвитию. Авторы статьи утверждают, что решения сотрудников, основанные на их навыках и сильно зависящие от их мотивации, теперь играют решающую роль в цепочке создания стоимости. Сотрудники в организации преследуют собственные цели, которые могут не совпадать с целями организации. Сотрудник может намеренно манипулировать информацией (когда считает это выгодным) и/или не выполнять поставленные планы и инструкции (опять же, если это сулит какую-то выгоду). Таким образом, сотрудник является источником «хаоса» в организационной структуре.

В статье исследователи предлагают подход, позволяющий системно рассматривать феномен активности сотрудников в организациях.

Авторы статьи [16], основываясь на своем многолетнем опыте как академических исследований, так и консалтинговых проектов, предлагают этапы распорядительной деятельности и координации по Файолю заменить на мотивацию, которая более важна для управления поведением сотрудников.

Анализ теоретических аспектов применения синергетической концепции управления в менеджменте на основе изучения периодической научной литературы позволяет сделать следующие выводы:

Во-первых, определено, что параметры порядка, стихийно возникающие в процессе преодоления кризисной ситуации или при прохождении индивидами точки бифуркации, координируют частные и общие интересы в конкретном сообществе и приобретают значение управленческих факторов, организующих движение. Организация проявляется как частный случай самоорганизации.

Во-вторых, анализируется возможность применения синергетических концепций социального управления для совершенствования механизмов управления ИРЦ РУП «Белтелеком». Рассматривается подход к идее использования энергии кризиса в социально-конструктивных целях при реализации проектов по разработке программного обеспечения информационных систем предприятия. Целью самоорганизации можно считать создание условий для синергетического эффекта во взаимодействии специалистов, участвующих в разработке.

Проработка этих идей в процессе разработки ИТ-проектов имеет перспективы дальнейшего исследования и разработки конкретных рекомендаций в конкретных условиях.

Направления совершенствования организационных механизмов управления ИРЦ РУП «Белтелеком»

Совершенствование организационного механизма управления процессом реализации заданий на разработку программного обеспечения ИС на основе автоматизации

Краткое описание внедренного механизма управления.

Автоматизирован механизм управления планированием, разработкой и тестированием прикладного программного обеспечения информационных систем предприятия в структурных подразделениях (цехах) ИРЦ с использованием системы Jira [12]. На ее основе организован учет проделанной работы программистами, координация взаимодействия при разработке ПО, контроль качества и сроков проделанной работы. План работ был разработан и принят к

реализации в рамках совещания «День оценки качества» производства ИРЦ. План реализовывался в виде еженедельных рабочих совещаний в рамках «День оценки качества».

Результат выполнения работы - внедрена в эксплуатацию информационная система Jira как инструмент управления проектами по разработке (изменению) программного обеспечения информационных систем для проектных команд на основе кросс-функциональных групп специалистов ЦАП, ЦРУЭС и ЦСиЭ производства ИРЦ. Схема организации процесса управления представлена на рисунке 6.

Эффект от внедрения механизма управления.

Автоматизация механизма управления разработкой ПО ИС значительно повысила его результативность. Положительный эффект сказался на всех этапах цикла управления: сбор информации, планирование, реализация, контроль, анализ, регулирование и координация.

Результат. В Jira ведется учет работ. Организация такого учета без автоматизации в вопросах координации деятельности трудновыполнима.

В ходе цикла управления выполнением задачи в секторе высока вероятность манипулирование агентом (инженером-программистом, технологом, проектировщиком) информацией о выполнении задачи, с целью снизить собственные затраты.



Рис. 6 – Совершенствование механизма управления процессом разработки ПО на основе автоматизации

Однако степень документированности процесса высокая, что позволяет более детально осуществлять планирования большого количества задач.

Совершенствование механизма оценки и контроля процесса реализации заданий на разработку программного обеспечения ИС на основе централизации данной функции управления

Краткое описание внедренного механизма управления.

Автоматизация механизмов управления разработкой ПО ИС дала возможность к совершенствованию механизма контроля за разработкой программного обеспечения.

В соответствии с инструкцией по разработке ПО ИС [11] считается выполненной и «закрытой», когда начальник ЦАП и начальник одного из цехов, ответственных за эксплуатацию ПО (ЦРУЭС, ЦСиЭ или один из филиалов) подписали акт о вводе ПО в опытно-промышленную или промышленную эксплуатацию. В ходе работы над актом программное обеспечение наполняется реальными данными и тестируется, устраняются все выявленные ошибки. В акте

фиксируется дата выполнения задания по разработке ПО ИС. Принятие программного обеспечения в эксплуатацию без оформления акта может повлечь в будущем ошибки при проведении, например, расчетов и быть причиной снижения коэффициента качества эксплуатации информационной системы, что недопустимо. Но в связи с автоматизацией остро встал вопрос к качеству оформления данных актов, а также к своевременности и качеству их внесения в автоматизированную систему Jira. Информация на основе актов является основой для документирования событий о выполнении задания, отчетности и планирования.

В группе производственного учета ИРЦ была введена должность специалиста. Функции обозначенные в управленческом цикле СПРУКАР как учет, контроль и анализ были выделены в отдельную подфункцию управления и в рамках должностных обязанностей поручены к выполнению специалисту. Данные функции предусматривают:

- учет результатов выполнения задач по разработке ПО ИС в системе Jira на основе утвержденных в установленном порядке [11] актов о вводе ПО в эксплуатацию;
- контроль – сравнение факта выполнения задачи с плановыми контрольными сроками;
- анализ – выявление совместно с исполнителями и изучение причин отклонения от контрольных сроков, подготовка отчетов для принятия решения.

Проведенные мероприятия позволили усовершенствовать механизм контроля за разработкой ПО ИС на основе централизации. Усовершенствованный механизм контроля представлен на рисунке 7.

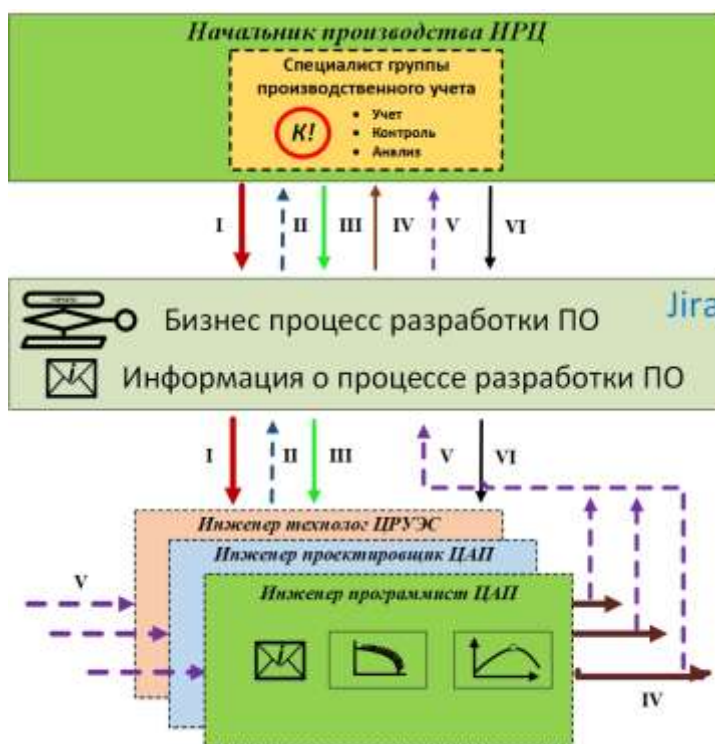


Рис. 7 – Совершенствование механизма контроля за разработки ПО

Эффект от внедрения механизма управления.

Установлен независимый от программистов, начальников секторов и цехов контроль за регистрацией в системе Jira выполнения задач, получения статистики по выполненным задачам, а также статистики по задачам, выполнение которых не укладывается в контрольные сроки. Информирование руководства ИРЦ о состоянии дел

Совершенствование механизма контроля создало условия для использования информации о нарушениях контрольных сроков при выполнении заданий в рамках еженедельных совещаний «День оценки качества» производства ИРЦ для оперативного контроля и принятия корректирующих действий.

Специалист группы производственного учета выполняет функции по ведению фонда алгоритмов и программ ИРЦ. В рамках данной функции он осуществляет регистрацию и учет

всей документации по проекту информационной системы, предусмотренной инструкцией [11]: заданий, технических заданий, актов ввода в эксплуатацию, инструкций пользователя к программному обеспечению и т.д.

Совершенствование механизма стимулирования в процессе управления реализацией заданий на разработку программного обеспечения ИС на основе механизма конкуренции

Краткое описание внедренного механизма управления.

Механизм мотивации описан в документе «Положение о мотивации работников цеха алгоритмов и программ производства «Информационно-расчетный центр» головного подразделения РУП «Белтелеком»» [14]. Структура оплаты труда показана на рисунке 8.

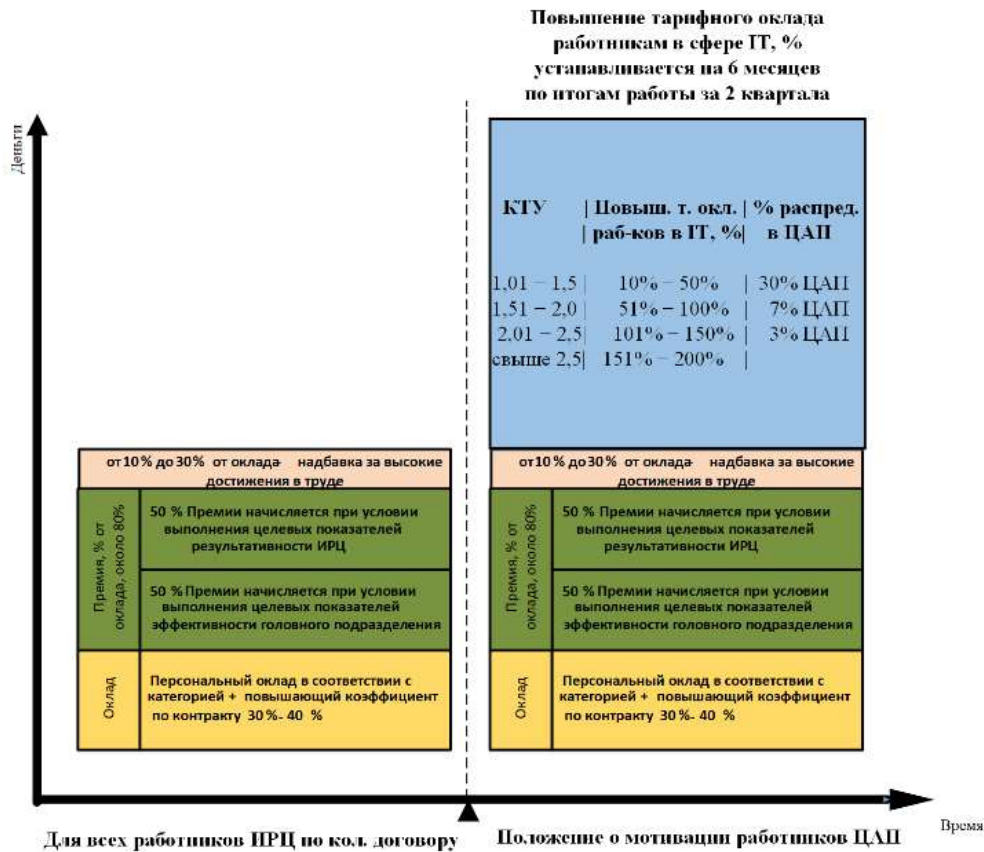


Рис. 8 – Структура оплаты труда инженеров-программистов ЦАП
Примечание – Источник: собственная разработка на основе [13, 14]

Определение персональных условий оплаты труда (повышение тарифного оклада высококвалифицированным работникам в сфере информационно-коммуникационных технологий) работникам ЦАП осуществляется в зависимости от качества их работы, своевременного выполнения проектов (задач), и их загруженности.

Эффект от внедрения механизма управления.

Значительно повысился уровень заинтересованности программистов в решении и поставленных задач. Особенно это касается опытных программистов, которые хорошо знают эксплуатируемые информационные системы и могут показать высокую результативность.

У руководства появилась возможность обратить внимание на программистов, работающих с низким КТУ. Ситуация с загруженностью программистов и их компетенцией стала прозрачной.

Заключение

В результате выполнения работы достигнута поставленная цель: разработаны и обоснованы рекомендации по повышению результативности системы управления разработкой ПО в ИРЦ РУП «Белтелеком». Решены все задачи исследования.

Основные результаты:

1. Изучены теоретические основы управления активными системами; показано, что ключевым ограничением действующей системы является игнорирование активности сотрудников, учет искажения ими информации, выбор состояния, не совпадающего с планом, недобросовестное поведение и другие действия.

2. Проведён анализ системы управления ИРЦ; выявлены «узкие места»: слабый контроль фактического выполнения задач, недостаточная мотивация опытных программистов, нецентрализованный учёт результатов.

3. Предложены и частично внедрены три механизма управления:

- автоматизация планирования и учёта на базе Jira;
- централизация функций контроля и анализа в группе производственного учёта;
- конкурентное стимулирование программистов через КТУ и дифференциацию окладов.

4. Показано, что внедрение этих механизмов повысило прозрачность, дисциплину и заинтересованность сотрудников.

Теоретическая значимость – работа демонстрирует применимость теории активных систем к реальному производственному подразделению ИТ-профиля и уточняет механизмы управления (контроль, стимулирование) для условий с высокой активностью агентов.

Практическая значимость – предложенные механизмы могут быть масштабированы на другие цеха ИРЦ и дочерние структуры «Белтелекома», а также использованы при автоматизации управления проектами.

Ограничения работы. Предложенные решения эффективны при относительно стабильной организационной структуре и достаточной квалификации персонала. В условиях высокой турбулентности или низкой ИТ-грамотности требуется адаптация.

Перспективы дальнейших исследований – внедрение Kanban/Scrum в Jira, использование Confluence для документации, интеграция с Bitbucket, развитие синергетического подхода (управление через самоорганизацию, использование энергии кризиса).

Так же перспективным является дальнейшее совершенствование механизмов конкурентного стимулирования [16], применяемого к программистам, что и реализует идею социального аттрактора и кредо развития современного социального управления: от принуждения к побуждению (т.е. мотивации) [1].

IMPROVING ORGANIZATIONAL CONTROL MECHANISMS FOR SOFTWARE DEVELOPMENT IN THE INFORMATION AND SETTLEMENT CENTER (ISC) OF RUE «BELTELECOM» BASED ON ACTIVE SYSTEMS THEORY

D.M. HARBUKOU

Abstract. The paper examines theoretical and practical aspects of software development management in the Information and Settlement Center (ISC) of RUE «Beltelecom» based on the theory of organizational systems management and the concept of agent activity. An analysis of existing management mechanisms (planning, motivation, control) is carried out, their drawbacks related to possible information distortion and deviations from plans are identified. Measures to improve organizational mechanisms are proposed and justified: automation of management based on Jira, centralization of accounting and control functions, introduction of competitive incentives for programmers. It is shown that the implementation of the proposed measures improves the efficiency of the ISC production management system. The possibilities of applying the synergetic approach and self-organization for further development of management mechanisms are also considered.

Keywords: management cycle, organizational systems management, active systems, control mechanisms, software development, RUE «Beltelecom» ISC, Jira, motivation, control, synergetics, self-organization.

Список литературы (применен стиль subtitle)

1. Касперович, Г. И. Синергетические концепции управления : курс лекций / Г. И. Касперович. – Минск : Акад. управления при Президенте Респ. Беларусь, 2005. – 258 с.
2. Руководство по качеству оказания услуг электросвязи РУП «Белтелеком» – Минск : РУП «Белтелеком», 2022. – 92 с.
3. Механизмы управления : учеб. пособие / В. Н. Бурков [и др.] ; под ред. Д. А. Новикова. – М. : УРСС (Editorial URSS), 2011. – 213 с.
4. Введение в теорию управления организационными системами : учеб. пособие / В. Н. Бурков [и др.] ; под ред. Д. А. Новикова. – М. : Книжный дом «Либроком»/ УРСС (Editorial URSS), 2009. – 264 с.
5. Новиков, А. М. Методология / А. М. Новиков, Д. А. Новиков. – М. : Синтег, 2007. – 668 с.
6. Официальный сайт РУП «Белтелеком» [Электронный ресурс]. – Режим доступа: <http://beltelecom.by/>. – Дата доступа: 11.10.2024.
7. Положение о производстве «Информационно-расчетный центр» РУП «Белтелеком» – Минск : РУП «Белтелеком», 2017. – 30 с.
8. Положение о цехе «Алгоритмов и программ» ИРЦ РУП «Белтелеком» – Минск : РУП «Белтелеком», 2016. – 16 с.
9. Положение о цехе «Расчетов за услуги электросвязи» ИРЦ РУП «Белтелеком» – Минск : РУП «Белтелеком», 2016. – 17 с.
10. Положение о цехе «Сопровождения и эксплуатации» ИРЦ РУП «Белтелеком» – Минск : РУП «Белтелеком», 2016. – 18 с.
11. Инструкция о подготовке и реализации заданий на разработку (изменение) программного обеспечения информационных систем РУП «Белтелеком» – Минск : РУП «Белтелеком», 2018. – 30 с.
12. Сайт компании ATlassian. Jira Software. Руководство по продукту и учебные материалы. [Электронный ресурс]. – Режим доступа: <https://www.atlassian.com/ru/software/jira/guides/getting-started/overview/>. – Дата доступа: 11.10.2024.
13. Коллективный договор РУП «Белтелеком» – Минск : РУП «Белтелеком», 2022. – 168 с.
14. Положение о мотивации работников цеха алгоритмов и программ производства «Информационно-расчетный центр» головного подразделения РУП «Белтелеком» – Минск : РУП «Белтелеком», 2022. – 9 с.
15. Положение об уполномоченном по качеству в структурном подразделении – Минск : РУП «Белтелеком», 2017. – 4 с.
16. Burkov, V. N. Mechanisms of Organizational Behavior Control: A Survey / V. N. Burkov, M. V. Goubko, N. A. Korgin, D. A. Novikov // *Advances in Systems Science and Applications*, Vol. 13, January 2013, № 1, pp. 1–13 [Electronic resource] – Mode of access: https://www.researchgate.net/publication/274390620_Mechanisms_of_Organizational_Behavior_Control_A_Survey - Date of access: 11.10.2024.
17. Mbengue, P. Management And Chaos Theory, Complexity Theory And Self-Organisation Systems Theory / Papa Mbengue, James Ondracek, M. Saeed, Andy Bertsch // *Asia Pacific Journal of Research in Business Management.*, Vol. 9, Issue 3, March 2018 Impact Factor: 5.16, ISSN: (2229-4104) [Electronic resource] – Mode of access: https://www.researchgate.net/publication/354309794_MANAGEMENT_AND_CHAOS_THEORY_COMPLEXITY_THEORY_AND_SELF-ORGANIZING_SYSTEMS_THEORY - Date of access: 11.10.2024.
18. Tschecher, W. Organization And Self-Organization / W. Tschecher, E. J. Brunner // *Springer Proceedings in Physics*, Vol.69, January 1992, Evolution of Dynamical Structures in Complex Systems, pp. 382-391 [Electronic resource] – Mode of access: https://www.researchgate.net/publication/303594014_Organization_and_Self-Organization - Date of access: 11.10.2024.
19. Five Functions of Management (Henri Fayol) [Electronic resource] – Mode of access: <https://intensescholar.com/management/five-functions-of-management-henri-fayol/#:~:text=Fayol's%20functions%20are%20Planning%2C%20Organizing,about%20the%20role%20of%20managers.> - Date of access: 11.10.2024.

ФОРМИРОВАНИЕ ЭТАЛОННЫХ ПАР СОВМЕЩЕННЫХ РАЗНОМОДАЛЬНЫХ РАЗНОВРЕМЕННЫХ ИЗОБРАЖЕНИЙ ДЛЯ ОЦЕНКИ КАЧЕСТВА ДЕТЕКТИРОВАНИЯ ЛИНИЙ

Е.М. ШЕМЕТОВИЧ, Г.А. ФРОЛОВ, М.И. ЗОРЬКО, В.В. ЧЕПИКОВА, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20.04.2026

Аннотация. Рассматривается задача оценки устойчивости линий на изображениях видимого и инфракрасного диапазонов к изменению погодных-временных условий видеорегистрации. Исследованы значения взвешенного количества несоответствующих объектов на эталонных парах, совмещенных разномодальных разновременных изображений. Установлено, что вейвлет-преобразование по сравнению с преобразованием Кэнни обеспечивает выделение более устойчивых линий, сохраняющих свое положение при изменении погодных-временных условий формирования изображений в видимом и инфракрасном диапазонах.

Ключевые слова: Гомография, разномодальные изображения, разновременные изображения, компьютерное зрение, OpenCV, Qt, совмещение изображений, детектор линий, вейвлет.

Введение

Для ориентирования в пространстве мобильный робот использует ключевые объекты. Часто в качестве таких объектов выступают точки изображений, локализуемые на изображениях в местах резкого изменения градиента с помощью соответствующих детекторов [1, 2]. В условиях открытого пространства изменение погодных-временных условий формирования изображений резко усложняет задачу сопоставления ключевых объектов изображений. Локализация точек в этом случае оказывается крайне нестабильной, что требует использования более устойчивых ключевых объектов, например, линий. Исследование устойчивости ключевых объектов к изменению погодных-временных условий становится критически важной задачей и требует создания набора эталонных данных для тестирования детекторов ключевых точек. Целью работы является разработка программного средства пространственного совмещения разномодальных изображений, необходимого для последующего анализа и обработки данных из различных источников (оптические, тепловизионные, мультиспектральные, рентгеновские изображения, данные радара с синтезированной апертурой).

Алгоритм полуавтоматического формирования эталонных пар, совмещенных разномодальных разновременных изображений

Разработан алгоритм полуавтоматического совмещения разномодальных разновременных изображений, полученных в общем случае разнодиапазонными сенсорами в различных масштабах, ракурсах, в различные моменты времени (различные сезонность, время суток), в разных погодных условиях. Алгоритм состоит из следующих шагов.

1. Загрузка исходных изображений $I_A(t_1)$ и $I_B(t_2)$, полученных камерами А и В в моменты времени t_1 и t_2 .

2. Назначение опорной точки $P_{ATL}(t_1)$ в верхней левой четверти изображения $I_A(t_1)$. Выбор для нее соответствующей опорной точки $P_{BTL}(t_2)$ изображения $I_B(t_2)$.

3. Назначение опорной точки $P_{ATR}(t_1)$ в верхней правой четверти изображения $I_A(t_1)$. Выбор для нее соответствующей опорной точки $P_{BTR}(t_2)$ изображения $I_B(t_2)$.

4. Назначение опорной точки $P_{ABL}(t_1)$ в нижней левой четверти изображения $I_A(t_1)$. Выбор для нее соответствующей опорной точки $P_{BBL}(t_2)$ изображения $I_B(t_2)$.

5. Назначение опорной точки $P_{ABR}(t_1)$ в нижней правой четверти изображения $I_A(t_1)$. Выбор для нее соответствующей опорной точки $P_{BRL}(t_2)$ изображения $I_B(t_2)$.

6. Формирование матрицы $H(I_A(t_1), I_B(t_2))$ гомографии [3] с помощью алгоритма f_{DTL} (DLT – Direct Linear Transformation) на основе сопоставления соответствующих пар опорных точек изображений $I_A(t_1)$ и $I_B(t_2)$, $H(I_A(t_1), I_B(t_2)) =$

$$= f_{DTL}(P_{ATR}(t_1), P_{BTR}(t_2); P_{ATR}(t_1), P_{BTR}(t_2); P_{ABL}(t_1), P_{BBL}(t_2); P_{ABR}(t_1), P_{BRL}(t_2)),$$

$$H(I_A(t_1), I_B(t_2)) = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}.$$

7. Гомография изображения $I_B(t_2)$. Формирование преобразованного изображения $I_B(t_2)$ в результате переноса каждого пикселя изображения $I_B(t_2)$ с координатами (y_B, x_B) в пиксель изображения $I_B(t_2)$ с координатами (y_B, x_B) , вычисляемыми с помощью выражений [3]

$$y_B = \frac{h_{21}x_B + h_{22}y_B + h_{23}}{h_{31}x_B + h_{32}y_B + h_{33}}, \quad (1)$$

$$x_B = \frac{h_{11}x_B + h_{12}y_B + h_{13}}{h_{31}x_B + h_{32}y_B + h_{33}}. \quad (2)$$

8. Обрезка изображений $I_A(t_1)$ и $I_B(t_2)$. Выбор двух точек на изображении $I_A(t_1)$, отмечающих вершины на диагонали прямоугольного окна, и поиск алгоритмами (ORB, AKAZE, BRISK) двух соответствующих точек на изображении $I_B(t_2)$. Фрагменты изображений $I_A(t_1)$ и $I_B(t_2)$, ограниченные заданным прямоугольным окном, сохраняются в качестве соответствующих эталонных изображений.

Программное обеспечение подготовки пар, совмещенных разномодальных разновременных изображений

Для реализации разработанного алгоритма полуавтоматического формирования пар совмещенных разномодальных разновременных изображений разработано соответствующее программное средство на языке C++ с использованием Qt и OpenCV. Программное средство обеспечивает загрузку пары изображений (рис. 1), интерактивный выбор на них опорных точек с использованием масштабируемых окон и пиксельной мозаики для повышения точности или альтернативно через ввод координат (рис. 2), с возможностью их последующего редактирования. После выбора соответствующих точек вычисляется матрица гомографии, осуществляется преобразование второго изображения относительно первого, визуализация наложения трансформированного изображения для верификации, предоставляется возможность выделения и сохранения прямоугольной области интереса с помощью соответствующих программных инструментов. Для упрощения работы оператора поддерживается возможность сохранения и загрузки данных об изображениях и уже выставленных на них точках. Программное средство

позволяет создавать высококачественные размеченные наборы данных, необходимые для обучения и верификации алгоритмов анализа.

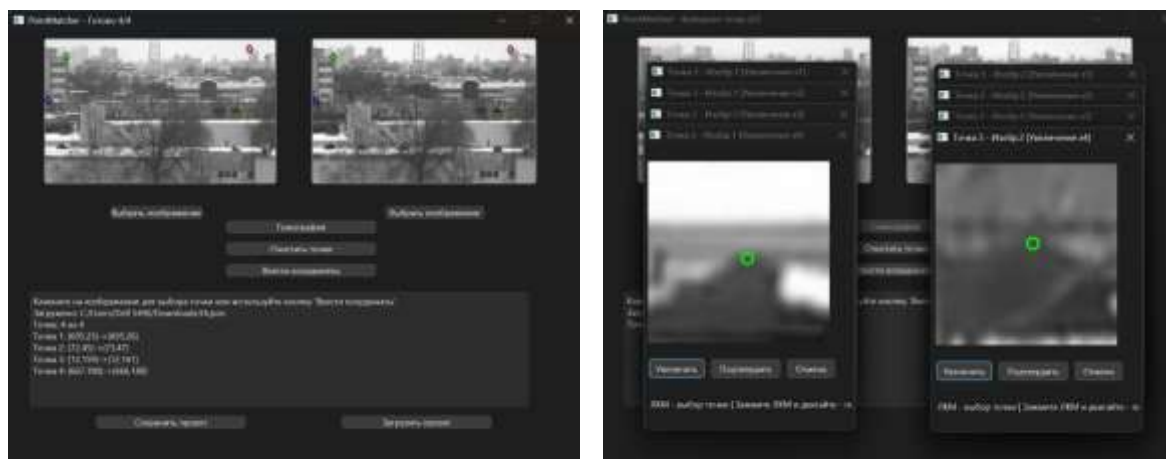


Рис. 1. Интерфейсы программы для загрузки изображений и выбора ключевых точек

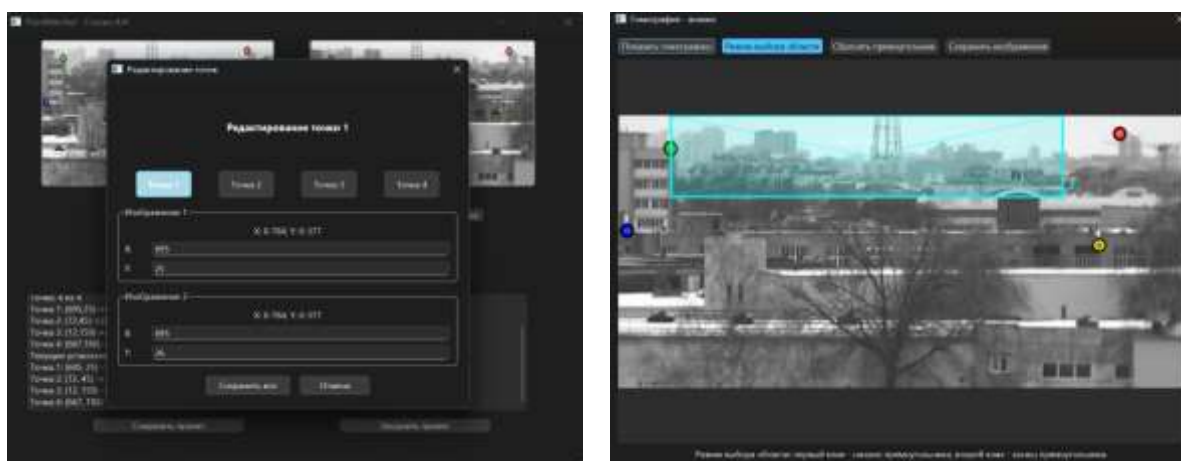


Рис. 2. Интерфейсы программы для ввода координат ключевых точек и формирования области интереса

Оценка качества детекторов линий с использованием эталонных пар совмещенных разномодальных разновременных изображений

Для оценки качества детекторов линий предлагается использовать устойчивость выделяемых ими линий к изменению погодных-временных условий формирования разномодальных изображений камерами видимого и инфракрасного диапазонов. Устойчивость линий предлагается оценивать по значению взвешенного количества N_E несоответствующих объектов, вычисляемые в результате приведения количества $N_E = N_E(t_1, t_2) + N_E(t_2, t_1)$ несоответствующих объектов $N_E(t_1, t_2)$ и $N_E(t_2, t_1)$ на изображениях $I_A(t_1)$ и $I_B(t_2)$ к общему количеству $N_O = N_O(t_1) + N_O(t_2)$ объектов на одном $N_O(t_1)$ и другом $N_O(t_2)$ изображениях с помощью выражения $N_E = \frac{N_E}{N_O}$. Изображения отличаются яркостью, контрастом, распределением теней, положениями динамических объектов. Различия в условиях формирования изображений приводят к изменению количества и положения линий, выделяемых детекторами. В результате для многих линий одного изображения отсутствуют линии в соответствующих координатах на другом изображении. В итоге, при сравнении разновременных кадров возникает значительное количество ложноположительных и ложноотрицательных результатов детектирования. На рис. 3 и 4 показаны эталонные пары совмещенных изображений

видимого и инфракрасного диапазонов снятые с разных ракурсов в разные периоды времени и комбинированные изображения, которые являются результатом их слияния.

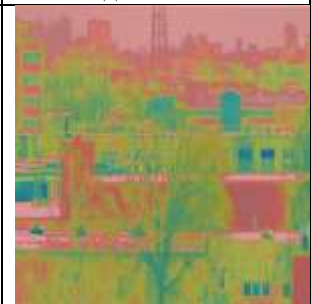
Изображения	Видимый диапазон	Инфракрасный диапазон	Комбинированный диапазон
$I_A(t_1)$			
$I_B(t_2)$			

Рис. 3. Эталонные пары совмещенных изображений (первый ракурс)

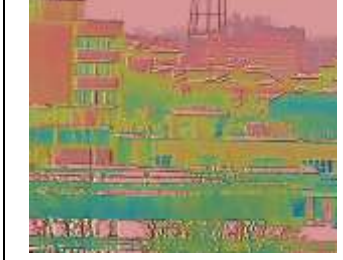
Изображения	Видимый диапазон	Инфракрасный диапазон	Комбинированный диапазон
$I(t_1)$			
$I(t_2)$			
$I(t_3)$			

Рис. 4. Эталонные пары совмещенных изображений (второй ракурс)

В табл. 1-3 приведены взвешенные ошибки (N_E) детектирования контуров детектором Кэнни и вейвлет-детектором, при коэффициенте перекрытия $k_x=0.75$ и дилатации $d=0$. Значение

$k_x=0.75$ означает, что 75 % линии на одном изображении совпало на втором, чтобы они считались совпавшими; $d=0$ означает, что контуры в экспериментах не расширяются.

Табл. 1. Взвешенные характеристики N_E результатов детектирования объектов (первый ракурс)

Детектор	Значения N_E для эталонных пар изображений в различных диапазонах		
	видимый	инфракрасный	комбинированный
Кэнни [3]	0,902	0,870	0,870
Вейвлет [4]	0,695	0,637	0,771

Табл. 2. Взвешенные характеристики N_E результатов детектирования объектов, полученные вейвлет-детектором (второй ракурс)

Сравниваемые пары изображений	Значения N_E для эталонных пар изображений в различных диапазонах		
	видимый	инфракрасный	комбинированный
$I(t_1)$ и $I(t_2)$	0,910	0,786	0,837
$I(t_1)$ и $I(t_3)$	0,764	0,729	0,723
$I(t_2)$ и $I(t_3)$	0,720	0,760	0,768

Табл. 3. Взвешенные характеристики N_E результатов детектирования объектов, полученные детектором Кэнни (второй ракурс)

Сравниваемые пары изображений	Значения N_E для эталонных пар изображений в различных диапазонах		
	видимый	инфракрасный	комбинированный
$I(t_1)$ и $I(t_2)$	0,982	0,978	0,983
$I(t_1)$ и $I(t_3)$	0,980	0,979	0,991
$I(t_2)$ и $I(t_3)$	0,984	0,976	0,988

Из табл. 1-3 следует, что линии, выделяемые с помощью вейвлет-преобразования, значительно более устойчивы к изменению погодных-временных условий формирования изображений по сравнению с линиями, выделяемыми с помощью преобразования Кэнни, для изображений всех диапазонов.

Оценка зависимости детектирования линий от показателей метрик оценивающих качество цифровых изображений

Оценка качества цифровых изображений (image quality assessment, IQA) – это неотъемлемый этап их анализа в различных автоматизированных системах. Понятие качества можно трактовать по-разному, в зависимости от целей использования изображений оно может быть связано как с визуальными, так и с художественными особенностями восприятия снимка человеком. Факторы, влияющие на качество, включают в себя равномерность освещения сцены и яркость изображения, его разрешение, контрастность, резкость и точность цветопередачи. Далее будут рассмотрены метрики $SSIM$ и корреляция Пирсона [5].

Функция вычисления индекса структурного подобия основана на локальных оценках сходства $ssim$, вычисленных между двумя окнами x и y размером $N \times N$ пикселей, рассчитанных вокруг соответствующих пикселей на изображениях A и B равных размеров согласно выражению:

$$ssim(x, y) = \frac{(2\mu_x\mu_y + c_1) * (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) * (\sigma_x^2 + \sigma_y^2 + c_2)},$$

где μ_x – среднее значение в окне x ; μ_y – среднее значение в окне y ; σ_x^2 – дисперсия значений в окне x ; σ_y^2 – дисперсия значений в окне y ; σ_{xy} – ковариация в окнах x и y ; константы $c_1 = (k_1L)^2$ и

$c_2=(k_2L)^2$, $L=255$, – значения динамического диапазона допустимых значений яркости; $k_1=0.01$, $k_2=0.03$ – экспериментально определенные константы. Итоговый индекс $SSIM$ для изображений A и B вычисляется как среднее арифметическое локальных оценок по формуле:

$$SSIM(A,B)=\frac{1}{M}\sum_{x,y}ssim(x,y),$$

где A и B – два изображения одинакового размера, M – количество окон. Каждая локальная оценка $ssim(x,y)$ рассчитывается в окрестности каждого пикселя с координатами (x,y) с некоторым шагом. Значения индекса $SSIM$ лежат в диапазоне $[-1,1]$.

Функция вычисления коэффициента корреляции Пирсона основана на анализе линейной зависимости между двумя изображениями A и B одинакового размера $N \times M$ пикселей. Она вычисляется как отношение ковариации значений яркости пикселей изображений к произведению их стандартных отклонений согласно выражению:

$$corr(A,B)=\frac{\sum_{i=1}^N\sum_{j=1}^M(I_A(i,j)-\bar{I}_A)(I_B(i,j)-\bar{I}_B)}{\sqrt{\sum_{i=1}^N\sum_{j=1}^M(I_A(i,j)-\bar{I}_A)^2}\cdot\sqrt{\sum_{i=1}^N\sum_{j=1}^M(I_B(i,j)-\bar{I}_B)^2}},$$

где $I_A(i,j)$ и $I_B(i,j)$ – значения яркости пикселей в координатах (i,j) изображения A и B соответственно; $\bar{I}_A$ и $\bar{I}_B$ – средние значения яркости по всему изображению A и B .

Значения коэффициента корреляции Пирсона лежат в диапазоне $[-1,1]$. Значение 1 достигается, когда изображения полностью идентичны (линейная зависимость); значение 0 указывает на отсутствие корреляции, а значение -1 свидетельствует о том, что изображения являются инвертированными (негативными) копиями друг друга. В отличие от индекса $SSIM$, коэффициент корреляции Пирсона вычисляется для всего изображения целиком (глобально), что обеспечивает его высокую вычислительную эффективность.

Табл. 4, рис. 5 являются сводными для показателей сходства изображений $SSIM$ и коэффициента корреляции Пирсона (*Pearson*). Из табл. 4 следует, что выбор метрик сходства влияет на оценку качества обработки: низкие показатели корреляции Пирсона в комбинированном диапазоне требуют использования более надежного структурного индекса $SSIM$.

Табл 4. Показатели структурного и линейного сходства изображений

Диапазон	Сравниваемые пары изображений	$SSIM$	$Pearson$	Сравниваемые пары изображений	$SSIM$	$Pearson$
Видимый	$I(t_1)$ и $I(t_2)$	0,71	0,85	$I(t_2)$ и $I(t_3)$	0,59	0,76
	$I(t_1)$ и $I(t_3)$	0,58	0,72	$I_A(t_1)$ и $I_B(t_2)$	0,81	0,94
Инфракрасный	$I(t_1)$ и $I(t_2)$	0,73	0,84	$I(t_2)$ и $I(t_3)$	0,69	0,78
	$I(t_1)$ и $I(t_3)$	0,69	0,73	$I_A(t_1)$ и $I_B(t_2)$	0,81	0,85
Комбинированный	$I(t_1)$ и $I(t_2)$	0,66	0,20	$I(t_2)$ и $I(t_3)$	0,63	0,42
	$I(t_1)$ и $I(t_3)$	0,60	0,17	$I_A(t_1)$ и $I_B(t_2)$	0,85	0,84

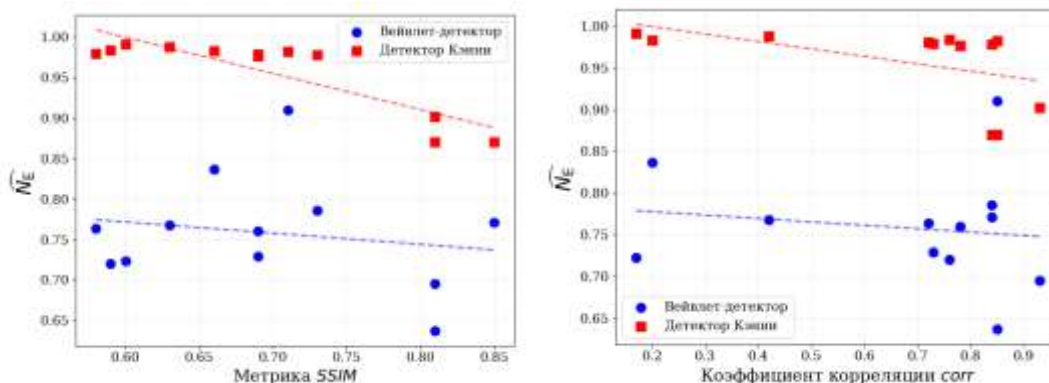


Рис. 5. Зависимости ошибок выделения контуров от $SSIM$ и коэффициента корреляции Пирсона

Из рис. 5 следует, что присутствует связь между сходством изображений, определяемых метриками и устойчивости детектирования линий.

Заключение

Разработан алгоритм формирования эталонных пар, совмещенных разномодальных разновременных изображений, позволяющий с помощью гомографии по четырем заданным парам опорных точек на двух совмещаемых изображениях получать их синхронизированные по растру фрагменты. Эталонные пары изображений использованы для оценки качества детекторов линий на основе взвешенного количества несоответствующих объектов. Установлено, что для изображений видимого и инфракрасного диапазонов вейвлет-преобразование по сравнению с преобразованием Кэнни обеспечивает выделение более устойчивых линий, сохраняющих свое положение при изменении погодных-временных условий формирования изображений. Установлена зависимость устойчивости детектирования линий к изменению погодных-временных условий их формирования от схожести изображений.

Благодарность. Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (БРФФИ) и Национального фонда естественных наук Китая (НФЕНК) в рамках проекта T25КИ-054 «Технология многокамерной локализации и построения карты открытого пространства».

FORMATION OF REFERENCE PAIRS OF MULTIMODAL MULTI-TEMPORAL IMAGES FOR EVALUATING LINE DETECTION QUALITY

E.M. SHEMETOVICH, G.A. FROLOV, M.I. ZORKO, V.V. CHEPIKOVA, V.YU. TSVIATKOU

Abstract. The problem of evaluating the stability of lines in visible and infrared range images to changes in weather and temporal video recording conditions is considered. The values of the weighted quantity of mismatched objects on reference pairs of aligned multimodal multi-temporal images are investigated. It is established that the wavelet transform, compared to the Canny transform, provides the extraction of more stable lines that maintain their position when weather and temporal conditions of image formation change in both visible and infrared ranges.

Keywords: Homography, multimodal images, multi-temporal images, computer vision, OpenCV, Qt, image alignment, line detector, wavelet.

Список литературы

1. Lowe D. // International Journal of Computer Vision. 2004. Vol. 60. P. 91–110.
2. Bay H., Tuytelaars T., Van Gool L. // Proc. of the 9th European Conference on Computer Vision (ECCV). 2006. P. 404–410.
3. Gonzalez R.C., Woods R.E. Digital Image Processing. 3rd ed. Prentice Hall, 2008.
4. Малла С. Вэйвлеты в обработке сигналов. М.: Мир, 2005.
5. Голуб Ю. И., Старовойтов В. В. Оценка качества цифровых изображений. Минск: ОИПИ НАН Беларуси, 2023.

MULTI-ROBOT TASK-PATH COLLABORATIVE OPTIMIZATION

XINYUE SHAO, QICHENG GUO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 09, 2026

Abstract. As robotics is widely used in intelligent manufacturing, logistics and unmanned systems, multi-robot systems have become a research hotspot for their high robustness, flexibility and parallelism. Task allocation and path planning are core to collaboration but face challenges in dynamic environments, communication limits and heterogeneous conflicts. This paper defines the problem, reviews architectures and algorithms, analyzes key challenges, and prospects future directions such as embodied intelligence and cloud-edge-device collaboration.

Keywords: Multi-robot systems, Task allocation, Path planning, Collaborative optimization.

Introduction

In major equipment manufacturing and service scenarios such as aerospace, smart warehousing, and urban inspection, the efficient processing of large-scale complex components and the real-time response to dynamic tasks impose extremely high demands on the efficiency, flexibility, and robustness of automation systems. Traditional manual or single-robot operation modes can no longer meet the demands of large-scale, high-precision operations. Multi-robot systems, leveraging their capabilities for parallel operation, functional complementarity, and strong fault tolerance, have become a key technological pathway to address these challenges [2]. However, simply stacking multiple robots together not only fails to improve efficiency but may instead lead to performance degradation due to behavioral conflicts and resource competition [3]. Therefore, designing an efficient decision-making core to achieve coordination among multiple robots is central to current research.

The core of multi-robot collaboration lies in solving two tightly coupled problems: task allocation and path planning. Task allocation aims to reasonably assign pending tasks to appropriate robots based on robot states and task requirements to optimize global objectives (e.g. minimizing total completion time, maximizing the number of completed tasks) [4]. Path planning generates collision-free trajectories for each robot from their start to target positions, ensuring the safety and feasibility of execution [5]. Early research often treated these as independent sub-problems, employing a sequential allocate-then-plan strategy. However, in complex dynamic scenarios, this decoupled approach can easily lead to task allocation that ignores path feasibility, or path planning constrained by rigid task assignments, resulting in suboptimal solutions or even deadlock [6].

In recent years, academia and industry have gradually recognized the deep coupling between task allocation and path planning, studying it as a joint optimization problem [7]. Concurrently, with the development of 6G communications, edge computing, and artificial intelligence, multi-robot systems are transitioning from structured static environments to unstructured dynamic ones, facing more severe challenges such as communication constraints, time-varying tasks, and heterogeneous collaboration. Thus, there is an urgent need to systematically review the collaborative technologies for multi-robot task allocation and path planning, clarifying the current state of research, key challenges, and development trends.

This paper aims to provide a comprehensive review of the collaborative optimization of task allocation and path planning in multi-robot systems. The structure of the paper is as follows: section 2 elaborates on the problem definition and system architectures, section 3 categorizes and reviews

mainstream algorithms, focusing on coupling strategies, section 4 analyzes the core challenges currently faced, section 5 concludes the paper.

Problem Definition and System Architectures

Problem Definition. The problem of multi-robot task allocation and path planning can be viewed as an extension and deepening of the classical Multi-Agent Path Finding (MAPF) problem. Classical MAPF aims to plan collision-free paths for multiple agents on a shared graph from start to goal positions, primarily focusing on path feasibility and optimality [5]. Cooperative Multi-Agent Path Finding (Co-MAPF), building upon this, introduces more complex collaboration mechanisms and integrates task allocation. Its core lies in enhancing overall system efficiency through active collaboration among agents (e.g., information sharing, goal coordination) while satisfying collision-free constraints [1].

In problem formulation, an undirected graph typically represents the workspace, along with a group of agents. Every agent has its own starting position, target position, and corresponding assigned task. The optimization objectives of Co-MAPF expand from minimizing the cost of a single path to the combined optimization of multiple indicators, including the number of finished tasks, total completion time, and system energy consumption. Compared with the traditional MAPF problem, Co-MAPF changes the way of collaboration from passive collision avoidance to active coordination. It also focuses on real-time information interaction and dynamic decision-making under the condition of information dependence.

System Architectures. Based on different decision-making mechanisms, multi-robot system architectures are primarily classified into three types: centralized, distributed, and hybrid. Their comparison is shown in Table 1.

Centralized architecture relies on a central control unit with global information, which formulates task allocation and path planning for all robots [3]. Its advantage lies in achieving globally optimal planning and high-quality paths. However, its disadvantages include exponentially increasing computational complexity with the number of robots, poor scalability, single point of failure risk, and extremely high demands on communication bandwidth and central node computing power [1]. This architecture is suitable for scenarios with a small number of robots and relatively static environments.

Distributed architecture features autonomous decision-making by each robot based on local perception and limited neighborhood communication, achieving coordination through negotiation or preset protocols [4]. Its advantages are strong robustness and good scaling capability, the failure of a single robot does not affect overall system operation. However, due to the lack of global information, decisions can easily fall into local optima, and the design of coordination protocols is complex [1]. This architecture is suitable for large-scale, highly dynamic scenarios, such as drone swarms.

Hybrid architecture aims to combine the advantages of both centralized and distributed approaches, often employing a hierarchical design [1]. An upper-layer centralized controller handles global task allocation and rough path guidance, while lower-layer robots perform distributed adjustments and obstacle avoidance based on real-time local information during execution. This architecture balances global optimality and local real-time responsiveness, but interface design and hierarchical communication coordination become new challenges [1].

Table 1. Comparison of Three System Architectures

Architecture Type	Advantages	Disadvantages	Applicable Scenarios
Centralized	Global optimization, high path quality	High computational load, single point of failure, poor scalability	Static, small-scale systems
Distributed	Strong robustness, good scalability, dynamic adaptation	Local optima, complex coordination protocols	Large-scale, dynamic swarms
Hybrid	Balances global optimization and local flexibility	Complex interfaces, hierarchical communication bottlenecks	Large-scale, complex tasks

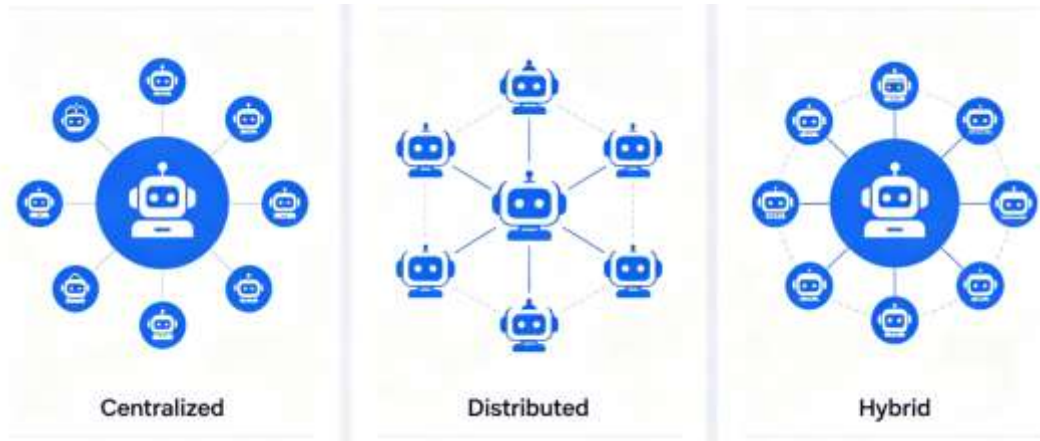


Figure 1. Three typical multi-robot system architectures

As shown in Figure 1, the three architectures differ significantly in decision-making mode and communication topology, which directly determines the performance of task allocation and path planning in multi-robot systems.

Collaborative Algorithms for Task Allocation and Path Planning

Based on algorithmic mechanisms, existing collaborative methods can be classified into search-based, sampling-based, intelligent optimization, learning-based, and communication-aware planning categories.

Search-Based Algorithms. Search-based algorithms abstract the environment into a graph and seek optimal solutions through systematic exploration. In Co-MAPF, typical methods include Hierarchical Cooperative A (HCA) and its improved version CO-WHCA, which dynamically identify and resolve path conflicts by introducing a conflict-oriented window mechanism. Conflict-Based Search (CBS) represents a state-of-the-art optimal solver, employing a high-level search to manage conflict constraints and a low-level search to plan paths for individual agents satisfying these constraints, improving efficiency while ensuring optimality. Furthermore, to address the coupling of task allocation and path planning, variants like TC-CBS have emerged, incorporating team coordination and multi-objective considerations into the search framework. Search-based algorithms excel in path optimality, but their computational complexity grows exponentially with the number of agents, posing real-time challenges in large-scale scenarios [1].

Sampling-Based Algorithms. Sampling-based algorithms, such as Rapidly-exploring Random Trees (RRT) and Probabilistic Roadmaps (PRM), construct paths by randomly sampling the state space, making them particularly suitable for handling high-dimensional and complex constraint problems [1]. In Co-MAPF, algorithms like Multi-RRT employ a decentralized iterative framework of plan-compare-assign treating conflicting parties as dynamic obstacles for local replanning, effectively addressing collaboration in shared environments. Improved PRM algorithms incorporate considerations like energy awareness and survival strategies to enhance robustness and energy efficiency in complex environments. A key advantage of such algorithms is their natural alignment with distributed architectures, supporting asynchronous parallel planning and reducing centralized computational load. However, their random sampling nature introduces uncertainty in path quality and convergence speed [1].

Intelligent Optimization Algorithms. Intelligent optimization algorithms, such as genetic algorithms, particle swarm optimization, and ant colony optimization, search for optimal or near-optimal solutions by simulating natural collective behaviors or evolutionary mechanisms, offering advantages in solving nonlinear, multi-objective optimization problems [3]. For example, the Decentralized Asynchronous Collaborative Genetic Algorithm (DAC-GA) balances local optimization and global collaboration through low-communication-overhead handover value exchanges. The hybrid algorithm combining PSO and Artificial Fish Swarm Algorithm (AFSA-PSO) leverages parallel group optimization and mutation mechanisms, balancing global convergence and local search capabilities. These algorithms can effectively handle the combinatorial explosion arising from the coupling of task

scheduling and path planning, but they typically incur high computational costs and require fine parameter tuning, making optimality guarantees difficult [3].

Learning-Based Algorithms. With the advancement of artificial intelligence, learning-based algorithms, particularly Deep Reinforcement Learning (DRL), have shown significant potential in addressing high-dimensional, dynamic multi-robot coordination problems.

Traditional Reinforcement Learning uses tabular methods like Q-learning with designed dynamic reward functions and collaboration priorities, suitable for discrete, static scenarios. However, it struggles with state-space explosion in complex environments [1].

Deep Reinforcement Learning approximates policies or value functions using neural networks, effectively solving high-dimensional continuous state-space problems. In Co-MAPF, the Centralized Training with Decentralized Execution (CTDE) framework, such as MADDPG, is widely used. It allows access to global information during training while relying only on local observations during execution, achieving efficient collaborative obstacle avoidance and path optimization. The introduction of Graph Neural Networks (GNNs) further enhances communication efficiency among multiple agents. GNNs inherently possess permutation equivariance, enabling them to handle dynamically changing communication topologies, allowing robots to aggregate information from neighboring nodes for more coordinated decisions. For instance, the MACNS framework integrates GNNs into policy networks to achieve homogeneous decision-making. Additionally, the PRIMAL series of algorithms combines imitation learning with reinforcement learning to learn implicit coordination strategies from local, avoiding explicit communication overhead. Although learning-based methods have high training costs, their powerful dynamic adaptability and generalization capabilities make them an ideal choice for handling complex, unknown scenarios [1].

Communication-Aware and Joint Optimization. In multi-robot systems, communication is not an ideal resource, its bandwidth, latency, and range constraints directly impact system performance. Therefore, collaborative planning that considers communication constraints has become an important research direction.

Communication-Aware Planning treats communication link constraints as part of the planning process, for example, planning robot paths to maintain connectivity with base stations or leader robots, or optimizing communication quality through formation control. Plan-Aware Communication conversely utilizes known planning information to optimize communication timing and content, such as scheduling high-priority communication during critical mission phases or transmitting only the most crucial state information under bandwidth limitations. Joint Planning simultaneously addresses path and communication problems through a unified optimization framework. For example, the Task-Control-Communication joint optimization model proposed in couples NOMA, beamforming, and trajectory planning, solved using hierarchical reinforcement learning, significantly improving task completion rates and communication speeds. This line of research signifies that multi-robot collaboration is evolving from a control-dominant approach towards an intelligent direction of deeply coupled control and communication.

Core Challenges

Despite significant progress in multi-robot task allocation and path planning, several challenges remain in practical applications.

Conflict between Algorithm Real-time Performance and Scalability. Search-based algorithms face exponential explosion when seeking optimal solutions, while heuristic or learning-based algorithms, though faster, struggle to guarantee global optimality. Achieving millisecond-level planning for hundreds or thousands of robots while maintaining solution quality remains a challenge [1].

Adaptation to Dynamic Environments and Uncertainty. Real-world scenarios involve dynamically changing obstacles, randomly arriving tasks, and time-varying communication links, demanding strong online replanning capabilities. Most existing methods rely on prior modeling and are insufficiently responsive to emergencies [3].

Collaborative Robustness under Communication Constraints. In practical deployment, communication inevitably suffers from delays, packet loss, and limited bandwidth. Traditional algorithms often assume ideal communication, ensuring the robustness of system coordination under constrained or even intermittent communication is crucial for practical application.

Deep Coupling between Task Allocation and Path Planning. Simple sequential processing struggles to achieve global optimality, while fully joint optimization faces extremely high solution complexity. Designing efficient coupling mechanisms that allow task allocation to perceive path costs and path planning to feedback task feasibility is key to improving system efficiency [7].

Coordination of Heterogeneity and Large-scale Swarms. Different types of robots (e.g., aerial, ground, manipulators) have varying kinematics, dynamics, and sensing capabilities, increasing the complexity of coordinated control. Simultaneously, the expansion of swarm size imposes higher demands on communication protocols and computing architectures [4].

Conclusion

Multi-robot systems are key enablers for future intelligent manufacturing and smart services. The collaborative optimization of task allocation and path planning is central to achieving efficient, robust, and intelligent operations. This paper provides a systematic review of the field from four perspectives: problem definition, system architectures, mainstream algorithms, and core challenges. The study indicates that task allocation and path planning are moving from separate design towards deeply coupled joint optimization, with communication constraints, dynamic adaptability, and scalability being the main current challenges. In the future, leveraging cutting-edge technologies such as embodied intelligence, graph neural networks, and cloud-edge-device collaboration to build an integrated communication-sensing-control collaborative framework will become an important direction in multi-robot system research, promising to unlock the significant potential of multi-robot collaboration in more complex real-world scenarios.

References

1. Xiong J, Zhang W, et al. // Journal of System Simulation. 2025.
2. D'Andrea R. // IEEE Transactions on Automation Science and Engineering. 2012.
3. Zhang Z, Mao J, et al. // Acta Automatica Sinica. 2024.
4. Ma H, Koenig S. // AAMAS 2016.
5. Stern R, Sturtevant N R, et al. // SoCS 2019.
6. Greshler N, Gordon O, et al. // IEEE. 2021.
7. Honig W, Kiesel S, et al. // AAMAS 2018.

ПРОЦЕДУРА ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ НА ОСНОВЕ МЕТОДА ИЗМЕНЕНИЯ СОСТАВА

А.М. ТИМОФЕЕВ¹, Д.В. ШЛЯХТУН²

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь,
2 – Национальный детский технопарк, Республика Беларусь

Поступила в редакцию 1 апреля 2026

Аннотация. Предложена структурная схема программной реализации алгоритма обезличивания данных на основе метода изменения состава, которая соответствует требованиям законодательства Республики Беларусь в сфере защиты информации. Процедура обезличивания включает в себя разделение исходных данных на 64-битные блоки и их последующее итерационное преобразование в соответствии с ГОСТ 28147-89 и характеризуется достаточно высоким уровнем информационной безопасности, так как значительно затрудняет статистический анализ со стороны нарушителя из-за изменения статистических свойств обезличенных персональных данных. Предложенная структурная схема процедуры обезличивания универсальна для любых атрибутов персональных данных и допускает реализацию при помощи программных, аппаратных или аппаратно-программных средств. Благодаря оптимизации алгоритма решение не налагает высоких требований к вычислительным мощностям пользовательского оборудования, обеспечивая высокую доступность технологии для информационных систем легитимных пользователей.

Ключевые слова: информационные системы, персональные данные, обезличивание персональных данных, методы обезличивания персональных данных, метод изменения состава или семантики.

Введение

Обеспечение конфиденциальности в современных информационных системах напрямую зависит от качественного обезличивания персональных данных [1 – 3]. Сегодня для этого активно применяют методы обезличивания персональных данных. Актуальность этой задачи в Республике Беларусь продиктована требованиями законодательства, которое обязывает операторов принимать меры по защите прав и свобод субъектов персональных данных при обработке их личных сведений [4, 5].

В соответствии с требованиями законодательства Республики Беларусь в сфере защиты информации одним из основных способов обезличивания персональных данных является метод изменения состава [4]. Суть этого метода заключается в том, что часть сведений о субъекте персональных данных обобщается, изменяется или удаляется. Однако этот подход имеет существенные недостатки. Он не обеспечивает достаточный уровень информационной безопасности при семантических преобразованиях исходных персональных данных, так как оставляет возможность их деобезличивания при помощи сопоставления с внешними источниками. Кроме того, удаление или изменение части данных заставляет оператора хранить исходные персональные данные отдельно для обеспечения их целостности и возможности восстановления. Это неизбежно увеличивает нагрузку на инфраструктуру и существенно увеличивает объем баз данных.

В настоящей работе представлена процедура обезличивания на основе метода изменения состава, лишенная описанных недостатков. Предложенный алгоритм не требует больших вычислительных ресурсов и подходит для оборудования легитимных пользователей. При этом предложенная процедура обезличивания позволяет существенно сократить объем хранимых

персональных данных по сравнению с существующими способами реализации метода изменения состава. Повышение уровня информационной безопасности достигается за счет трансформации статистических свойств данных, что исключает возможность их успешного анализа злоумышленником.

Потому целью исследования стала разработка такой процедуры обезличивания, которая повысит уровень информационной безопасности обезличенных персональных данных и оптимизирует ресурсы системы. Объектом работы стал метод изменения состава. Предметом исследования являлось повышение уровня информационной безопасности обезличенных персональных данных благодаря изменению их вероятностных характеристик.

Процедура обезличивания персональных данных

На рис. 1 представлена структурная схема, описывающая процедуру обезличивания персональных данных на основе метода изменения состава или семантики.

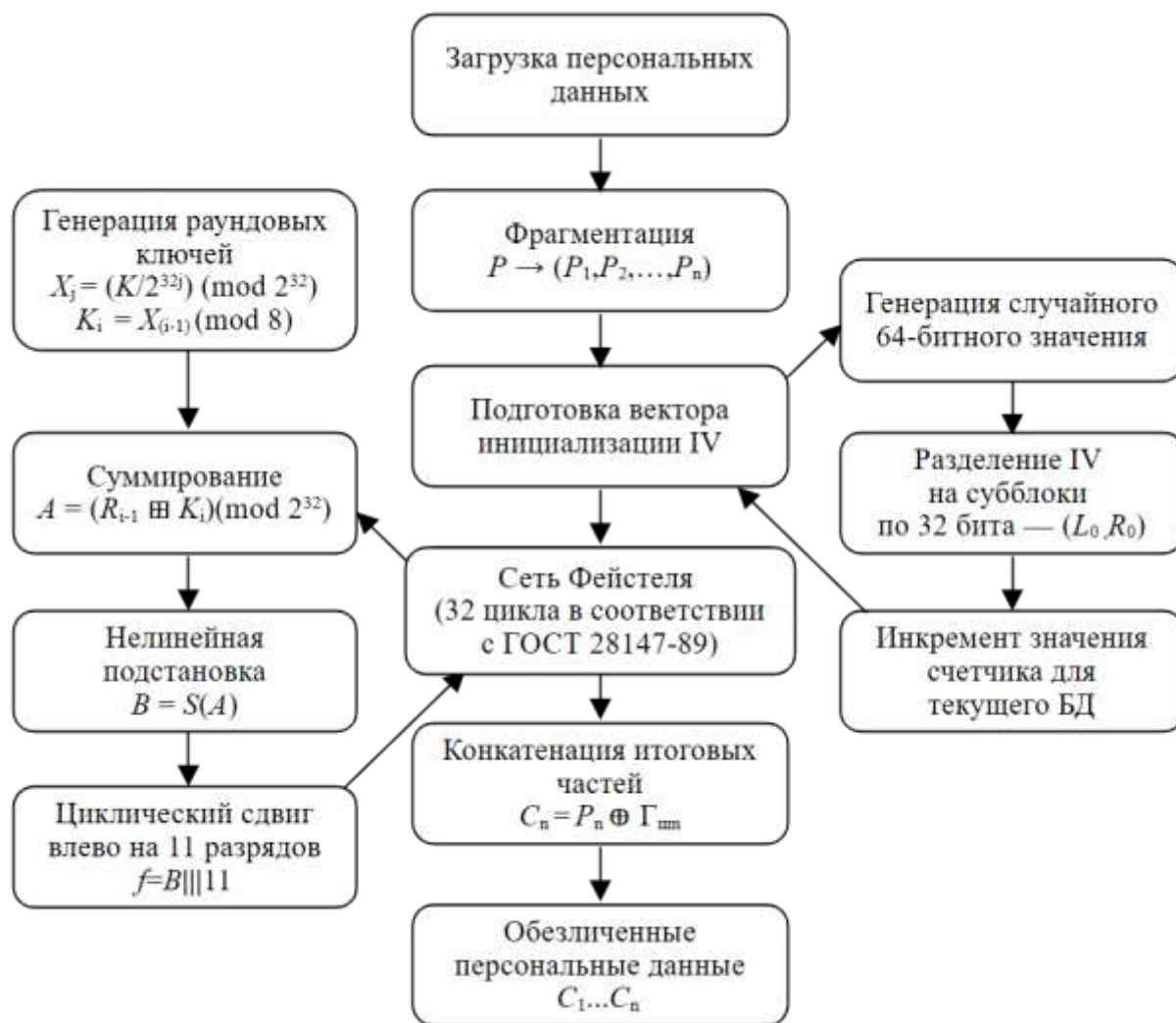


Рис. 1. Структурная схема обезличивания персональных данных, построенная на основе метода изменения состава

Процедура обезличивания персональных данных в соответствии с данной структурной схемой представляет собой итерационный программный алгоритм, который преобразует данные при помощи их посимвольного перевода в кодировку UTF-16, разбиения на 64-битные блоки и 32-циклового шифрования с использованием вектора инициализации и нелинейных подстановок [6]. Результатом процесса является конкатенация преобразованных блоков и их обратное преобразование в символьный вид.

Заключение

Разработана структурная схема программной реализации алгоритма обезличивания персональных данных на основе метода изменения состава, которая полностью соответствует требованиям законодательства Республики Беларусь в сфере защиты информации.

Применение итерационного процесса на основе ГОСТ 28147-89 с использованием изменяющегося вектора инициализации и процедур нелинейной подстановки позволяет изменить структуру исходных данных. Статистический анализ показывает, что энтропия полученных обезличенных персональных данных значительно выше, чем для исходных. Это сводит к минимуму вероятность распознавания содержания обезличенных персональных данных злоумышленником при помощи методов частотного или статистического анализа из-за изменения их статистических свойств.

Эффективность предложенного решения дополнительно повышается за счет внедрения предложенного метода фрагментации, который позволяет существенно сократить объем обрабатываемой информации. Переход от стандартного 16-битного представления UTF-16 к использованию 9-битного формата (управляющий флаг и 8 бит значимой части кода) обеспечивает значительную экономию дискового пространства и оперативной памяти. Такой подход гарантирует уменьшение объема данных при сохранении их полноты и возможности корректного восстановления исходного текста легитимными пользователями.

PERSONAL DATA DEPERSONALIZATION PROCEDURE BASED ON THE COMPOSITION CHANGE METHOD

A.M. TIMOFEEV, D.V. SHLIAKHTUN

Abstract. A structural diagram of the software implementation of the data depersonalization algorithm based on the composition change method is proposed, which complies with the requirements of the legislation of the Republic of Belarus in the field of information security. The anonymization procedure involves dividing the source data into 64-bit blocks and then iteratively transforming them in accordance with GOST 28147-89. This procedure offers a high level of information security, as it significantly complicates statistical analysis by an attacker due to changes in the statistical properties of the anonymized personal data. The proposed structural diagram of the anonymization procedure is universal for any personal data attributes and can be implemented using software, hardware, or a combination of hardware and software. Thanks to algorithm optimization, the solution does not place high demands on the computing power of user equipment, ensuring high availability of the technology for the information systems of legitimate users.

Keywords: information systems, personal data, depersonalization of personal data, methods of depersonalization of personal data, method of changing the composition or semantics.

Список литературы

1. Ворона В.А. Биометрическая идентификация личности. М., 2023.
2. Коллинз М. Защита сетей. Подход на основе анализа данных. М., 2020.
3. Остапенко Г.А. Информационные операции и атаки в социотехнических системах: организационно-правовые аспекты противодействия. М., 2020.
4. Приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 10 декабря 2024 г. № 259 «Об изменении приказов Оперативно-аналитического центра при Президенте Республики Беларусь от 28 марта 2014 г. № 26 и от 20 февраля 2020 г. № 66» [Электронный ресурс]. – Режим доступа: <https://www.oac.gov.by/public/content/files/files/law/prikaz-oac/2024%20-%20259.pdf>. – Дата доступа: 19.03.2026 г.
5. Закон Республики Беларусь от 7 мая 2021 г. № 99-3 «О защите персональных данных» [Электронный ресурс]. – Режим доступа: <https://pravo.by/document/?guid=12551&p0=N12100099>. – Дата доступа: 19.03.2026 г.
6. ГОСТ 28147-89. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200007350>. – Дата доступа: 19.03.2026.

MULTI SENSOR FUSION FOR MOBILE ROBOT PERCEPTION

BIN ZHANG

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received April 06, 2026*

Abstract. This paper proposes a multi-sensor fusion system using Light Detection and Ranging (LiDAR), Inertial Measurement Unit (IMU), RGB-Depth (Red Green Blue-Depth) camera, and ultrasonic sensors to overcome single-line LiDAR limitations in robot navigation. By integrating a tightly-coupled LiDAR-IMU odometry and vision-based 3D mapping, the design significantly improves positioning accuracy and obstacle awareness. The distributed architecture ensures robust performance for autonomous mobile platforms.

Keywords: 2D LiDAR, IMU, RGB-D Camera, ROS2, SLAM, Obstacle Avoidance.

Introduction

With the rapid development of artificial intelligence and sensor technologies, mobile robots have gradually expanded from industrial automation to daily life scenarios such as healthcare, logistics, and domestic service [1]. Whether it is automated guided vehicles in factories, delivery robots in hotels, or vacuum cleaning robots in home environments, all require the ability to operate safely and reliably in unknown or dynamic environments. LiDAR, regarded as the eyes of mobile robots, has become a core sensor for autonomous navigation due to its high ranging accuracy and fast response speed [2]. Within the ecosystem of the open-source robot operating system Robot Operating System 2 (ROS2) [3], the development of low-cost mobile robot platforms has become a current research hotspot.

Single-line LiDAR, with its advantages of high measurement accuracy, good real-time performance, and controllable cost, has become one of the core sensors in mobile robot perception. However, when the system relies solely on a single LiDAR sensor to simultaneously undertake both obstacle avoidance and environmental mapping tasks, it faces several unavoidable theoretical limitations and engineering challenges. Firstly, in terms of perception dimensionality, single-line LiDAR can only obtain two-dimensional planar information and cannot perceive obstacle height, leading to the risk of missed detection of low-height or suspended objects. Additionally, the near-field blind zone limits the response space for close-range collision avoidance. Secondly, at the level of localization and mapping, pure LiDAR odometry lacks the motion priors provided by wheel encoders or IMU, making it prone to pose mutation due to scan matching failure during rapid rotation or movement. Furthermore, single-line LiDAR has perception blind zones for special materials such as glass and mirrors, struggles to distinguish between static obstacles and moving objects in dynamic environments, and the entire perception system suffers from the vulnerability of a single point of failure.

To validate the aforementioned limitations, this study proposes a mobile robot platform featuring a distributed control system with cooperative upper and lower computers (see Figure 1). In terms of system architecture, the upper computer utilizes a Raspberry Pi 4B running the Ubuntu operating system and the ROS2 Humble framework, primarily responsible for LiDAR data acquisition, obstacle avoidance algorithm computation, and visualization of point cloud data. The lower computer employs an Arduino Uno, communicating with the upper computer via Universal Serial Bus (USB) serial port, receiving speed control commands, performing inverse kinematics calculations for the Mecanum wheels, and ultimately controlling four Direct Current (DC) motors through an L293D motor driver module. This division of labor between the upper and lower computers ensures both the operational efficiency of complex algorithms and the real-time performance of low-level control, providing a reliable hardware foundation for subsequent obstacle avoidance and mapping experiments. It decouples

complex algorithm processing from real-time low-level control, fully leveraging the performance advantages of different processors.

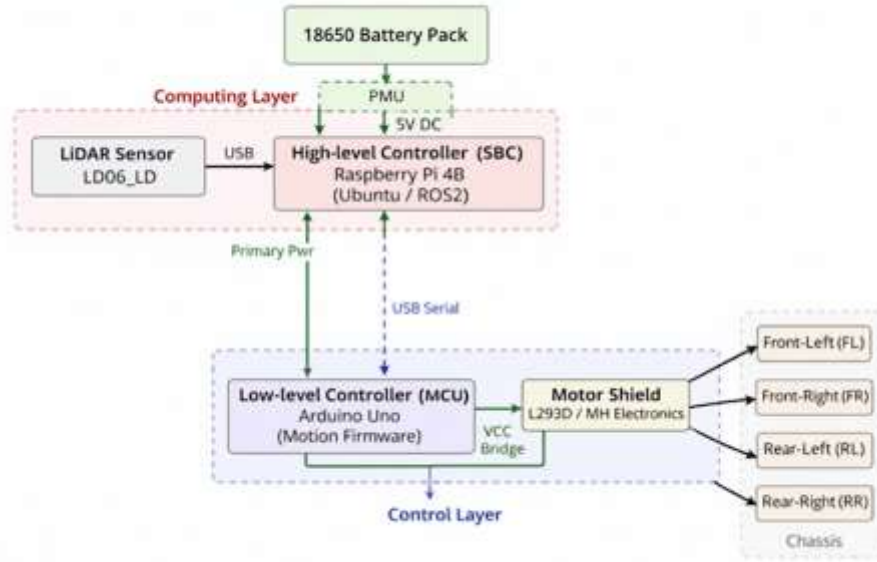


Figure 1. System hardware architecture diagram

In preliminary Simultaneous Localization and Mapping (SLAM) mapping and obstacle avoidance experiments, the three theoretical issues mentioned above were fully validated in the real system. The experimental phenomena are specifically manifested in the following three aspects:

First, Mapping Drift Caused by Motion Estimation Degradation. Due to the system not being equipped with an IMU or high-precision wheel encoders, the pure LiDAR odometry exhibited significant mapping drift during rapid rotation and movement, specifically manifested as point cloud ghosting and pose offset (see Figure 2). Furthermore, the single LiDAR sensor has an inherent defect during the mapping process: when dynamic objects appear in the environment, the system cannot distinguish them from static obstacles, leading to anomalies or even crashes in the mapping results.

Second, Obstacle Missed Detection Caused by Spatial Perception Limitations. The LD06 2D LiDAR is mounted on the robot's roof (approximately 10 cm above the ground) and can only acquire two-dimensional scan data within the horizontal plane, resulting in significantly insufficient three-dimensional spatial perception capability. Although the radar demonstrates good obstacle avoidance response to suddenly appearing obstacles within the horizontal plane, when the obstacle height is below the LiDAR scan plane, the system fails to detect it entirely, leading to direct collisions.

Third, Lack of Environmental Adaptability and System Robustness. When the robot approaches glass windows or mirrored furniture, severe missing occurs in the LiDAR point cloud over large reflective surfaces, causing local localization loss. Moreover, the system relies entirely on LiDAR as the sole perception source. When the sensor experiences temporary disconnection due to loose interfaces or other reasons, the robot immediately falls into a "perception failure" state, unable to continue mapping or perform obstacle avoidance tasks.

Addressing the three problems above, this paper proposes corresponding solutions, which will be elaborated in subsequent sections: First, to address spatial perception limitations, an RGB-D camera will be introduced to obtain obstacle height information, compensating for the perception blind zones outside the LiDAR plane. Second, to tackle motion estimation degradation, an IMU will be fused with LiDAR to construct a tightly-coupled odometry, suppressing pose drift during rapid motion and on rough terrain. Finally, to improve environmental adaptability and system robustness, fault detection and degraded operation mechanisms will be added to the obstacle avoidance strategy, along with discussions on handling dynamic objects.

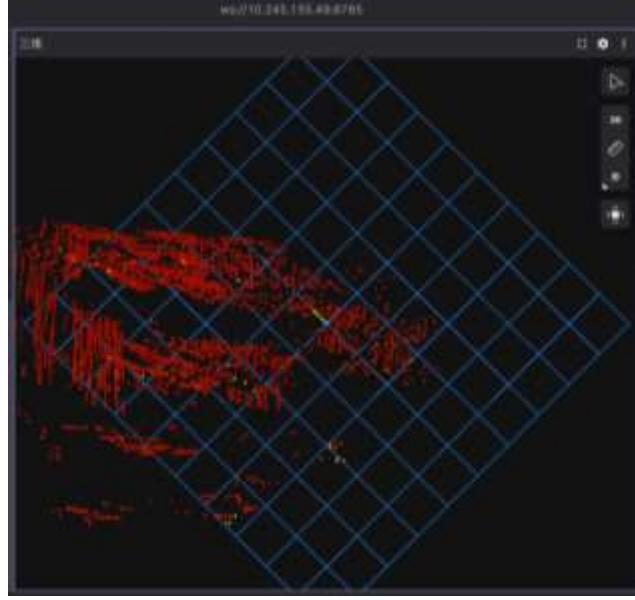


Figure 2. Mapping drift

Suppression of Motion Estimation Degradation: Design of LiDAR-IMU Fusion Odometry

Addressing the pose drift problem of pure LiDAR odometry during rapid rotation, in feature-sparse environments, and on rough terrain identified in the experiments, this paper proposes integrating an IMU with LiDAR based on the structure shown in Figure 1 to construct a tightly-coupled LiDAR-IMU odometry. An MPU6050 six-axis IMU is selected, with an output frequency of 200 Hz. Time alignment with the LiDAR is achieved using ROS2's joint time synchronization mechanism. The fusion algorithm adopts a tightly-coupled framework based on an Error State Kalman Filter [4]: IMU data provides high-frequency state propagation for pose prediction between two LiDAR scans; upon arrival of a new LiDAR scan, feature points are extracted and scan-matched against the local map to obtain observation residuals; the filter update corrects the accumulated IMU error and estimates IMU biases online. In feature-sparse environments, the system adaptively increases the weight of the IMU prediction to avoid pose mutation caused by scan matching failure [4]. To address changes in vehicle pitch angle on rough terrain, the attitude angles estimated by the IMU are fed back to the LiDAR point cloud preprocessing module. The raw point cloud is rotationally corrected before scan matching, effectively suppressing mapping distortion caused by vehicle tilt. This fusion odometry provides a stable pose estimation foundation for the entire system, supporting the subsequent implementation of mapping and obstacle avoidance functions.

Enhancement of System Robustness: 3D Environmental Perception Based on an RGB-D Camera

Building upon the stable pose estimation provided by the LiDAR-IMU fusion odometry, this paper introduces a RGB camera as the primary mapping sensor to address the vertical perception blind zone problem of single-line LiDAR. The LD06 LiDAR can only acquire two-dimensional horizontal information, failing to detect low-height or suspended obstacles and unable to identify obstacle categories. Therefore, the system defaults to using the RGB camera as the core mapping sensor, running an Oriented Fast and Rotated Brief SLAM2 (ORB-SLAM2) [5] visual-inertial tightly-coupled SLAM system. In environments with good lighting and rich texture, it constructs a 3D point cloud map containing obstacle height information and, combined with lightweight object detection algorithms, identifies obstacle types that LiDAR cannot perceive.

Ensuring seamless connect of maps constructed by the two different sensors, the system does not support simultaneous operation of LiDAR and the RGB camera. Consequently, a manual switching mode is adopted in practical deployment: By default, the RGB camera is responsible for mapping and object recognition. When the robot enters scenes with insufficient lighting, texture scarcity, or excessive

dynamic objects, leading to a decline in visual SLAM tracking quality, the user can manually shut down ORB-SLAM2 via upper computer commands and switch to LiDAR SLAM mode. In this mode, the system relies on the LiDAR-IMU fusion odometry to maintain mapping and localization stability. The pose information provided by this odometry is used to align the coordinate systems of the maps, ensuring seamless connection of maps.

Regarding obstacle avoidance functionality, regardless of which sensor is currently used for mapping, the LiDAR always remains active as the primary obstacle avoidance sensor. Leveraging its high ranging accuracy and fast response speed, LiDAR ensures real-time obstacle avoidance performance for the robot in most scenarios. Simultaneously, the position information of low-height and suspended obstacles identified by the RGB camera during mapping is synchronized to the obstacle avoidance module. When the robot approaches these areas, the system triggers supplementary obstacle avoidance responses, forming a dual mechanism with LiDAR as the primary and the RGB camera as the secondary sensor. This design prioritizes the RGB camera for mapping, fully utilizing its advantages in 3D perception and object recognition, while relying on LiDAR for real-time obstacle avoidance, achieving a balance between perception completeness and operational safety under limited computational resources.

Enhancement of System Robustness: Near-Field Obstacle Avoidance and Fault Tolerance Using Ultrasonic Sensors

At the obstacle avoidance functional level, LiDAR, due to its high ranging accuracy and fast response speed, is always activated as the primary obstacle avoidance sensor. It is responsible for real-time detection of and response to obstacles within the horizontal plane, ensuring operational safety in most scenarios. However, 2D LiDAR has inherent limitations: its scanning plane is fixed, rendering it completely ineffective against transparent or reflective materials like glass and mirrors, and it possesses a near-field blind zone. Mounted approximately 10 cm above the ground on the robot's top, areas above and below its scanning plane, particularly low-height obstacles and uneven pavement, are entirely outside its detection range.

To compensate for these defects, this section introduces an RGB camera and ultrasonic sensors as supplementary obstacle avoidance methods. In its default mapping mode, the RGB camera continuously identifies low-height obstacles, suspended obstacles, and uneven pavement (ground height changes) using visual SLAM and object detection algorithms, recording the location information of these hazardous areas in the map. When the robot subsequently approaches these pre-identified areas, the obstacle avoidance module proactively triggers detour or deceleration responses. However, the RGB camera may fail under conditions of insufficient lighting, texture scarcity, or excessive dynamic objects, rendering it unable to identify newly appearing obstacles in real-time.

Two to three HC-SR04 ultrasonic sensors are installed on the front of the robot chassis to provide gapless scanning of the area directly in front of the robot, focusing on detecting low-height obstacles below the LiDAR's scanning plane. Ultrasonic sensors are insensitive to transparent and reflective materials, effectively detecting obstacles like glass and mirrors that LiDAR struggles with. In the obstacle avoidance logic, LiDAR always holds the first priority, responsible for real-time detection of the vast majority of obstacles, with its safety threshold set at 30 cm. The ultrasonic sensors act as a second line of defense, configured with a more conservative safety threshold of 15 cm. When LiDAR fails to trigger avoidance due to special materials, blind zones, or height limitations, and an ultrasonic sensor detects an obstacle within 15 cm, the system immediately executes deceleration and backward turning to re-plan the path. The RGB camera handles the predictive avoidance of pre-identified dangerous areas, complementing the real-time detection capability of the ultrasonic sensors: the camera addresses the avoidance of known hazardous areas, while the ultrasonic sensors handle real-time detection of sudden transparent obstacles and low-height obstructions.

Conclusion

This chapter addresses the inherent limitations of single-line LiDAR by introducing an IMU, an RGB camera, and ultrasonic sensors, thereby constructing a multi-sensor fusion perception and localization system. The LiDAR-IMU fusion odometry mitigates pose drift, providing a stable pose

foundation for the system. The RGB camera serves as the default mapping sensor, building 3D point cloud maps via ORB-SLAM2 to solve the vertical perception blind zone problem. The ultrasonic sensors act as supplementary obstacle avoidance tools, filling gaps related to transparent materials and the near-field blind zone. The three sensors have distinct roles, collectively enhancing the system's perceptual completeness and operational safety.

Due to the performance limitations of the hardware platform, this study has the following shortcomings: The computational load on the Raspberry Pi is high, potentially leading to lag during prolonged operation; limited battery life restricts the ability to support large-scale mapping tasks; cumulative drift remains unavoidable during long-duration mapping; the system lacks the ability to detect uneven pavement (ground height changes) in dark environments where the RGB-D camera fails; and the effectiveness and reliability of the fault-tolerant degraded mode require further experimental validation. Given the constraints of the aforementioned hardware platform and sensory capabilities, the experiments in this research phase will primarily be conducted in structured indoor environments, including home settings, laboratories, and corridors, to validate the basic functionality and cooperative mechanisms of the multi-sensor fusion framework. Adaptability testing in outdoor environments and complex unstructured scenarios is planned as future research direction.

References

1. Islamgozhayev T., Kalimoldayev M., et al. // Open Eng. 2016. Vol. 6. P. 347–352.
2. Zhao Y., Li Y., Shang Y., et al. // Telemetry & Telecontrol. 2014. No. 5. P. 4–22.
3. Macenski S., Foote T., et al. // Science Robotics. 2022. Vol. 7. P. eabm6074.
4. Xu W., Zhang F. // IEEE Robotics and Automation Letters. 2021. Vol. 6. P. 3317–3324.
5. Mur-Artal R., Tardós J. D. // IEEE Transactions on Robotics. 2017. Vol. 33. P. 1255–1262.

DYNAMIC CHANNEL MODELING AND PERFORMANCE ANALYSIS FOR UAV COMMUNICATIONS

CHI ZHANG

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 10, 2026

Abstract. This paper presents a comprehensive dynamic channel modeling and simulation framework for unmanned aerial vehicle (UAV) air-to-ground communication systems. Starting from fundamental AWGN and Rayleigh fading channels, we progressively incorporate large-scale path loss (distance-dependent) and Doppler shift (velocity-dependent) to construct a realistic time-varying channel model. The performances of BPSK and QPSK modulations are compared under both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios. Simulation results quantitatively reveal the impact of flight distance and velocity on bit error rate (BER), and provide optimal modulation switching thresholds under different SNR conditions. An interactive GUI tool is developed to visualize the dynamic BER variation and constellation diagrams in real time. The work offers theoretical insights and practical guidelines for UAV communication system design.

Keywords: UAV communications, channel modeling, path loss, Doppler shift, Rayleigh fading, BER analysis.

Introduction

Unmanned Aerial Vehicles (UAVs) have become essential in various applications such as surveillance, delivery, and emergency communications [1]. Reliable air-to-ground communication is critical for these missions, as it enables real-time data transmission and command control [2]. Unlike terrestrial communication links, UAV channels exhibit high dynamics due to platform mobility, altitude variations, and complex environmental interactions [3]. Accurate and comprehensive channel modeling is therefore fundamental for the design, simulation, and performance evaluation of robust UAV communication systems [4].

Existing studies have investigated UAV channel characteristics through field measurements and statistical modeling approaches [5-7]. However, most prior works focus on either static scenarios or simplified path loss models, lacking a systematic simulation framework that integrates both large-scale fading (path loss) and small-scale fading (Rayleigh fading) with real-time UAV trajectory dynamics [8]. Moreover, the joint effect of flight distance and velocity on modulation performance, particularly under varying line-of-sight (LOS) and non-line-of-sight (NLOS) conditions, is often treated separately rather than analyzed as a coupled system [9].

In this paper, we develop a step-by-step dynamic channel simulation framework that progresses from a basic additive white Gaussian noise (AWGN) channel to a complete time-varying model incorporating distance-dependent path loss, velocity-dependent Doppler shift, and Rayleigh fading. We compare the performance of binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK) modulations under both LOS and NLOS scenarios, derive quantitative relationships between flight parameters and bit error rate (BER), and propose adaptive modulation switching thresholds based on channel conditions. An interactive graphical user interface (GUI) is implemented to facilitate real-time parameter adjustment, dynamic visualization of BER curves, and constellation diagram display. The work provides both theoretical validation and practical guidelines for adaptive modulation and link budget design in UAV communication systems.

System Model

We consider a UAV flying horizontally at a constant altitude $h=100\text{m}$ with constant velocity $v=20\text{m/s}$ along the x-axis. The ground station is fixed at the origin. The UAV starts at $(0,0,100)$ and moves to $(1000,0,100)$. The instantaneous distance between UAV and ground station is

$$d(t) = \sqrt{(x(t))^2 + h^2}, \quad (1)$$

where $x(t) = vt$ is the horizontal displacement of the UAV at time t ; $v=20\text{m/s}$ is the constant flight velocity of the UAV.

The free-space path loss in dB is given by

$$PL(d) = 20\log_{10}(d) + 20\log_{10}(fc) - 147.55, \quad (2)$$

where $fc = 2.4\text{Hz}$ is the carrier frequency. The linear power gain is $g_{PL}(t) = 10^{-PL(d(t))/10}$.

The radial velocity component is

$$v_r(t) = \frac{x(t)v}{d(t)}. \quad (3)$$

The Doppler frequency shift is

$$f_d(t) = \frac{v_r(t)}{\lambda}, \lambda = \frac{c}{f_c}, \quad (4)$$

where $c = 3 \times 10^8 \text{ m/s}$. The Doppler-induced phase rotation is incorporated as $e^{j2\pi f_d(t)t}$.

For NLOS scenarios, we adopt a flat Rayleigh fading model with independent realizations per symbol

$$h_{ray}(t) = \frac{1}{\sqrt{2}}(N(0,1) + jN(0,1)), \quad (5)$$

ensuring $E[|h_{ray}|^2] = 1$, where $\alpha(t)$ follows Rayleigh distribution and $\theta(t)$ follows uniform distribution in $[0, 2\pi]$.

Two modulation schemes are considered: BPSK and QPSK. Both are normalized to unit average power. The transmitted symbols are generated from random bits. The received signal after the channel is

$$r(t) = \sqrt{g_{PL}(t)}h(t)e^{j2\pi f_d(t)t}s(t) + n(t), \quad (6)$$

where $h(t) = 1$ for LOS, and $h(t) = h_{ray}(t)$ for NLOS, and $n(t)$ is AWGN with power spectral density N_0 . The noise power is fixed based on a reference SNR at distance 100 m.

Simulation

Key simulation parameters are listed in Table 1.

Table 1. Simulation parameters

Parameter	Value
Carrier frequency	2.4 GHz
UAV velocity	20 m/s
UAV altitude	100 m
Flight distance range	0 – 1000 m
Symbol rate	1 kHz
Reference distance	100 m
Reference SNR range	0 – 30 dB
Number of symbols	50,000

A graphical user interface (GUI) is built using MATLAB App Designer, as shown in Figure 1. The GUI allows users to select modulation (BPSK/QPSK), enable/disable Rayleigh fading, adjust reference SNR via a slider, and instantly view the constellation diagram and BER-vs-distance curve. This tool facilitates intuitive understanding of how flight parameters affect communication performance.

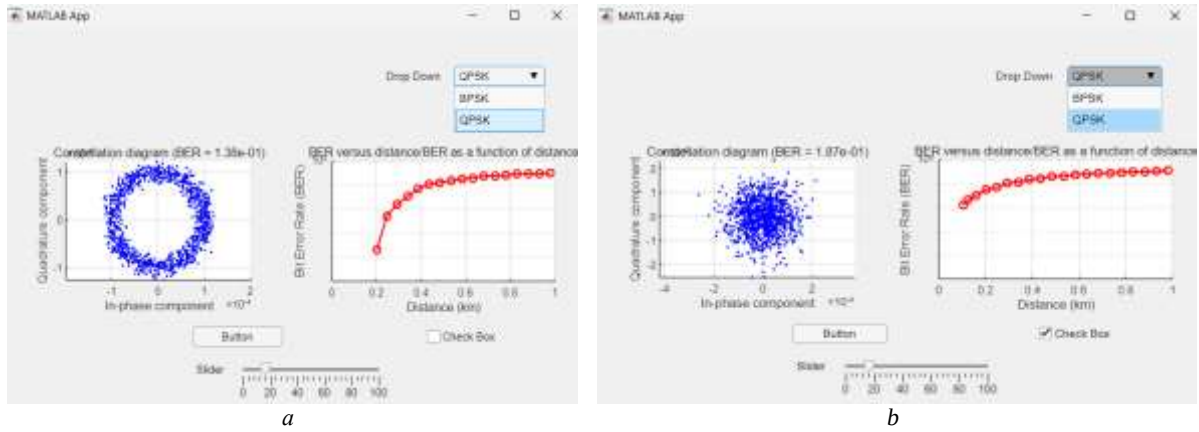


Figure 1. GUI for UAV channel simulation: a – close Check Box; b – open Check Box

Figure 2 shows the BER as a function of flight distance for BPSK and QPSK under LOS conditions (reference SNR = 30 dB at 100 m). As the UAV moves away, path loss increases, reducing received SNR and elevating BER. BPSK maintains lower BER than QPSK at all distances; at 1 km, BPSK BER is about 1.5×10^{-2} , while QPSK BER exceeds 10^{-1} . This indicates that QPSK is more sensitive to distance due to its higher SNR requirement.

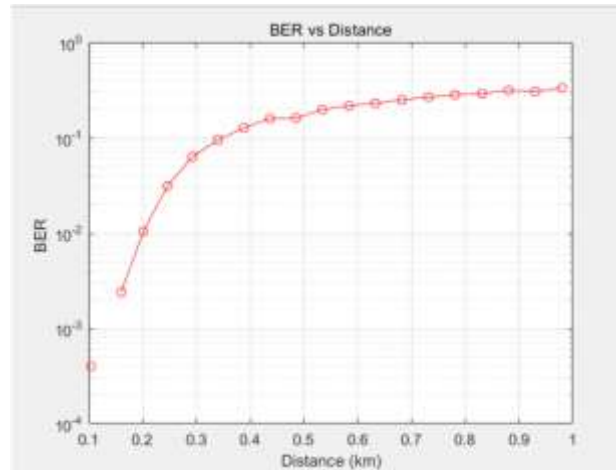


Figure 2. BER versus distance for BPSK and QPSK (LOS, reference SNR = 30 dB)

To isolate velocity effect, we fix the distance at 500 m and vary UAV speed from 10 to 50 m/s, with reference SNR = 20 dB. Figure 3 illustrates BER growth with increasing speed, especially for QPSK. The Doppler shift causes phase rotation that challenges coherent detection; without ideal channel tracking, higher speeds degrade performance more severely.

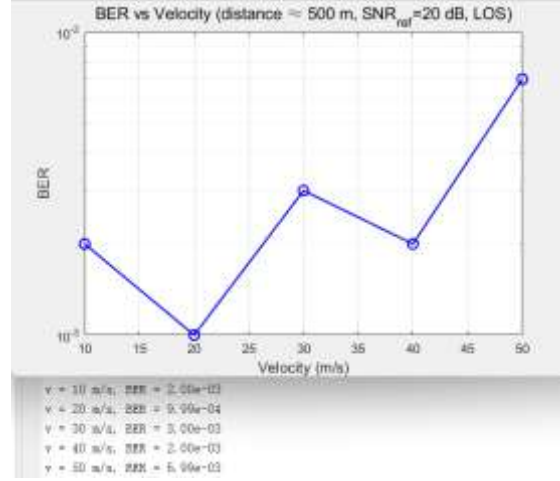


Figure 3. Situations that cannot be measured under a binocular system

Figure 4 compares BPSK and QPSK BER versus reference SNR under both LOS and NLOS (Rayleigh) channels. In LOS, BPSK achieves BER 10^{-3} to 10^{-4} at about 7 dB, while QPSK requires 10 dB. In NLOS, both suffer severe degradation: BPSK needs 18 dB, QPSK needs 22 dB for the same BER. This leads to the following modulation selection guideline:

1. $\text{SNR} < 10$ dB: use BPSK (reliability priority).
2. $10 \text{ dB} \leq \text{SNR} < 22$ dB: use QPSK (higher data rate while meeting BER target).
3. $\text{SNR} \geq 22$ dB: higher-order modulations (e.g., 16QAM) may be considered.

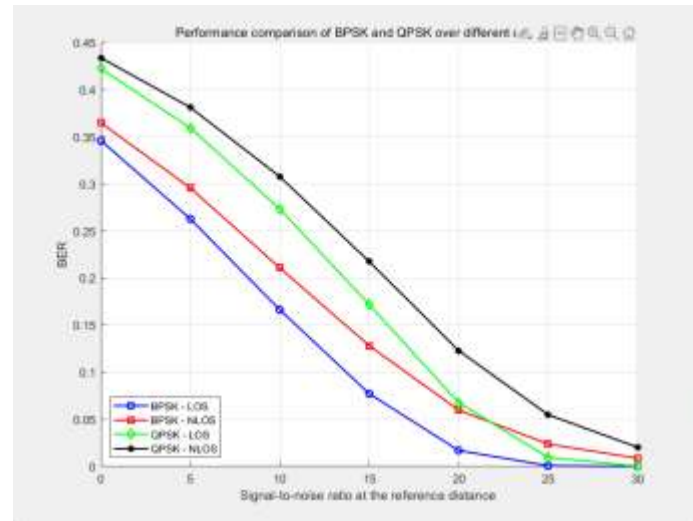


Figure 4. BER versus reference SNR for BPSK and QPSK under LOS and NLOS channels

Conclusion

This paper presented a comprehensive dynamic channel modeling and simulation framework for UAV communications. By progressively incorporating path loss, Doppler shift, and Rayleigh fading, we systematically analyzed the impact of flight distance and velocity on BER for BPSK and QPSK modulations. Key findings include: Distance significantly degrades BER, with QPSK being more sensitive than BPSK. Velocity-induced Doppler shift causes performance loss, especially for QPSK. SNR thresholds for modulation switching are identified: BPSK below 10 dB, QPSK between 10 and 22 dB.

An interactive GUI tool was developed to visualize these effects in real time, serving both educational and engineering design purposes. The work provides theoretical validation and practical insights for adaptive modulation and link budget in UAV communication systems. Future work may extend to more complex environments (e.g., urban canyon) and advanced modulations (e.g., OFDM).

References

1. Khawaja W., Guvenc I., Matolak D. W., Fiebig U. C., Schneckenburger N. // IEEE Communications Surveys & Tutorials. 2019. Vol. 21. No. 3. P. 2361-2391.
2. Matolak D. W., Sun R. // IEEE Transactions on Vehicular Technology. 2017. Vol. 66. No. 1. P. 26-44.
3. Lin X., Yajnanarayana V., Muruganathan S. D., et al. // IEEE Communications Magazine. 2018. Vol. 56. No. 4. P. 204-210.
4. Simon M. K., Alouini M.-S. // Digital Communication over Fading Channels. 2nd ed. Wiley, 2005.
5. 3GPP // TR 36.777. 2017.
6. Mozaffari M., Saad W., Bennis M., Nam Y.-H., Debbah M. // IEEE Communications Magazine. 2017. Vol. 55. No. 5. P. 106-112.
7. Chen J., Liu L., Zhang Y., Xu Z. // IEEE Transactions on Antennas and Propagation. 2018. Vol. 66. No. 10. P. 5427-5438.
8. Zhang S., Li Y., Wang H. // IEEE Internet of Things Journal. 2021. Vol. 8. No. 12. P. 9745-9756.
9. Li Y., Zhou Z., Yang L. // IEEE International Conference on Communications Workshops (ICC Workshops). 2020. P. 1-6.

ANALYSIS OF CYCLE CONSISTENT GENERATIVE ADVERSARIAL NETWORKS FOR UNPAIRED IMAGE TRANSLATION

YU LIU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 31, 2026

Abstract. Unpaired image translation has become an important research topic in computer vision due to the difficulty of obtaining paired datasets in real-world scenarios. Cycle consistent generative adversarial networks (CycleGAN) provide an effective solution by learning mappings between two image domains without paired supervision. This paper presents a comprehensive analysis of the CycleGAN model, including its architecture, training strategy, and key components such as cycle consistency and identity mapping mechanisms. Experiments are conducted on benchmark datasets, including horse-to-zebra and seasonal translation tasks, and evaluated using SSIM and PSNR metrics. The results demonstrate that CycleGAN can generate visually realistic images while effectively preserving the structural content and essential features of the input images. Compared with traditional supervised methods, it significantly reduces the reliance on labeled data while still achieving competitive performance in various scenarios. However, limitations such as insufficient semantic consistency, visual artifacts in complex regions, and limited controllability are also discussed in detail, indicating directions for future improvement.

Keywords: CycleGAN, generative adversarial networks, image translation, deep learning, computer vision.

Introduction

Image-to-image translation is a basic task in computer vision. The goal is to transform an image from one domain into another while keeping the main content unchanged. This problem appears in many real situations, such as changing weather conditions, converting day images to night images, or translating medical images between different types.

Traditional methods depend on paired datasets, where each input image has a corresponding output image. However, such datasets are difficult to collect. In many cases, it is impossible to obtain two images of exactly the same scene under different conditions. This limits the use of supervised learning methods.

Generative adversarial networks have improved image generation by introducing a generator and a discriminator [1]. The generator tries to create realistic images, while the discriminator tries to detect fake images. Through this process, the model learns the data distribution.

However, early models still require paired data. CycleGAN was proposed to solve this problem and can work with unpaired datasets.

CycleGAN Architecture

CycleGAN is built on a dual mapping framework that consists of two generators and two discriminators, enabling bidirectional translation between two image domains. One generator G learns a mapping from domain X to domain Y , while the other generator F performs the inverse mapping from domain Y back to domain X . This bidirectional design allows the model to learn cross-domain relationships without requiring paired training data, which significantly increases its applicability in real world scenarios.

Each domain is associated with a corresponding discriminator. The discriminator DX distinguishes between real images from domain X and generated images produced by F , while DY performs the same task for domain Y . The discriminators are trained to classify real and fake samples, whereas the generators are optimized to produce images that are indistinguishable from real ones. This adversarial learning process enables the model to approximate the underlying data distributions of both domains.

The generators are typically implemented using convolutional neural networks with residual learning mechanisms. A standard generator consists of an encoder, several residual blocks, and a decoder. The encoder extracts high-level feature representations from the input image, the residual blocks transform these representations while preserving important structural information, and the decoder reconstructs the translated image. This design helps maintain the global structure of the input while performing domain transformation.

Residual blocks play a critical role in improving model performance. By learning residual mappings instead of direct transformations, they simplify the optimization process and help preserve essential image content. This is particularly important in image translation tasks where structural consistency must be maintained.

Instance normalization is commonly used in the generator networks. It normalizes feature statistics for each image independently, which improves training stability and enhances the model's ability to generalize across different data distributions.

The discriminators are usually implemented as Patch Generative Adversarial Network (PatchGAN) classifiers. Instead of evaluating the entire image, PatchGAN operates on local patches, allowing it to focus on high-frequency details such as textures and edges. As a result, the generated images exhibit improved local realism and finer visual details.

A key component of CycleGAN is the cycle consistency mechanism. When an image is translated from domain X to domain Y using generator G and then mapped back to domain X using generator F , the reconstructed image should be close to the original input. This constraint enforces consistency between forward and backward mappings, prevents arbitrary transformations, and ensures that the learned mappings preserve the semantic content of the images [2].

Training Strategy

The training process of CycleGAN is based on alternating optimization between generators and discriminators. In each iteration, the generators first produce translated images, and then the discriminators evaluate whether these images are real or generated. This process continues until both parts reach a balance.

Another important technique is identity mapping. When an image from the target domain is input into the generator, the output should remain close to the input. This helps preserve color and brightness information.

In practice, the Adam optimizer is commonly used. Learning rate scheduling is also applied to improve convergence. Proper parameter tuning is important because this model can be unstable during training.

Experimental Results

To evaluate the performance of the CycleGAN model, experiments were conducted on several public datasets. The horse-to-zebra dataset is one of the most commonly used benchmarks. In this task, the model learns to convert horses into zebras while preserving the overall shape and structure of the original images.

Another dataset involves seasonal translation, where images are converted between summer and winter scenes. This task is more challenging because it requires significant changes in lighting conditions, color distribution, and texture details, while still maintaining the underlying scene structure.

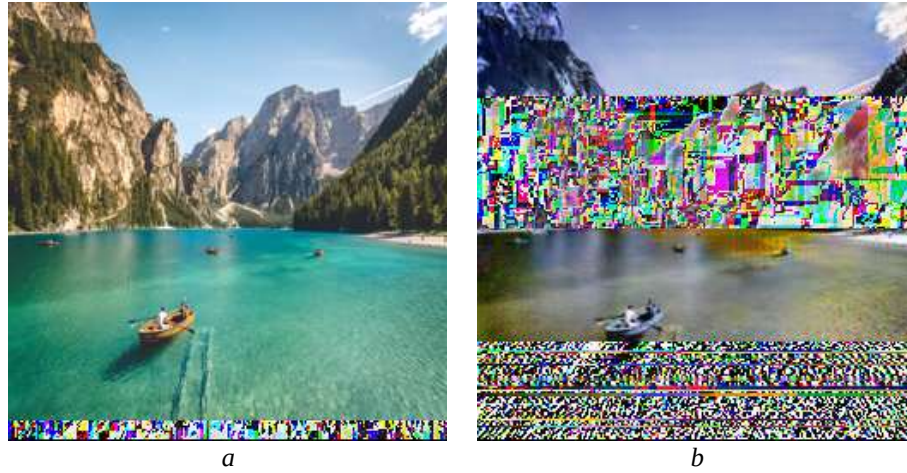


Figure 1. Example of seasonal image translation: *a* – summer; *b* – winter

As shown in Figure 1, the model effectively converts the summer image into a winter scene while preserving the overall structure of the original image.

Two evaluation metrics were used: structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). SSIM measures the similarity between images in terms of structure, while PSNR evaluates the overall reconstruction quality.

The experimental results show that CycleGAN is capable of generating visually realistic images while effectively preserving the main structural content of the input. Although its performance is slightly lower than that of supervised methods in some cases, it eliminates the need for paired training data, which represents a significant advantage in practical applications [3].

Applications and Limitations

CycleGAN has been widely used in various domains due to its ability to learn mappings from unpaired data. In medical imaging, it can be applied to translate between different imaging modalities such as MRI and CT, which improves data utilization and reduces the need for paired datasets. In computer graphics, CycleGAN enables style transfer tasks, such as converting photographs into artistic images. In autonomous driving, it can be used to enhance simulated data so that it better matches real-world environments, thereby improving model generalization.

The experimental results in this paper, including the horse-to-zebra and seasonal translation tasks, show that CycleGAN can generate visually realistic images while preserving the overall structure of the input. These findings are consistent with previous work, which demonstrates that the cycle consistency constraint helps preserve content during translation.

However, several limitations remain. First, when the difference between the two domains is large, the model may fail to preserve important semantic structures. This issue has been discussed in previous studies, where the mapping is learned without explicit semantic supervision [4].

Second, the model may produce visual artifacts, especially in regions with complex textures or fine details. This is a common limitation of adversarial training, where the objective focuses more on visual realism than structural accuracy.

In addition, CycleGAN provides limited control over the translation process. The mapping between domains is learned implicitly, making it difficult to control specific attributes or generate diverse outputs. This limitation has motivated later work on multimodal image translation [5] and contrastive learning-based approaches [6].

Overall, while CycleGAN provides a flexible solution for unpaired image-to-image translation and reduces the reliance on labeled data, its performance is still limited by the lack of semantic constraints and controllability. Addressing these issues remains an important direction for future research.

Conclusion

This paper presents an overview of image-to-image translation methods, with a focus on CycleGAN. Compared with supervised approaches such as Pix to Pix, CycleGAN can be trained using unpaired data, making it more flexible for real-world applications. The comparison results show that CycleGAN achieves competitive performance while reducing the need for labeled datasets.

However, as discussed in the previous section, CycleGAN still has several limitations. In particular, it may fail to preserve semantic consistency when the domain gap is large, and it may introduce visual artifacts in complex scenes. In addition, the lack of controllability limits its applicability in tasks that require precise transformations.

Future work can focus on improving semantic consistency and controllability. Possible directions include incorporating attention mechanisms, introducing additional constraints, or exploring more advanced models for image translation.

Overall, CycleGAN provides an effective solution for unpaired image-to-image translation, but further improvements are needed to enhance its robustness and applicability.

References

1. Goodfellow I., Pouget-Abadie J., Mirza M., et al. // Proc. of the Neural Information Processing Systems. 2014. P. 2672–2680.
2. Zhu J., et al. // Proc. of the IEEE International Conference on Computer Vision. 2017. P. 2223–2232.
3. Isola P., et al. // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 1125–1134.
4. Taigman Y., et al. // Proc. of the International Conference on Learning Representations. 2017.
5. Huang X., et al. // Proc. of the European Conference on Computer Vision. 2018. P. 172–189.
6. Park T., et al. // Proc. of the European Conference on Computer Vision. 2020. P. 319–345.

SEMANTIC COMMUNICATION SYSTEM FOR VEHICLE DETECTION

QIKAI WANG

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received April 08, 2026*

Abstract. This paper presents a semantic communication system for vehicle detection over noisy wireless channels. A Convolutional Neural Network (CNN) semantic encoder compresses image features with multiple compression ratios to transmit task-related information only, which is then processed by a pre-trained You Only Look Once version 12 (YOLOv12) object detector. A Supernet with shared encoder weights and Sandwich Rule training is proposed to reduce redundancy. Experiments on Additive White Gaussian Noise (AWGN) channels show that the semantic system significantly outperforms traditional transmission at low signal-to-noise ratio (SNR), and the Supernet achieves comparable performance with independent models.

Keywords: semantic communication, vehicle detection, YOLOv12, CNN encoder, supernet, AWGN channel, compression ratio.

Introduction

The rapid proliferation of intelligent transportation systems (ITS) has driven growing demand for reliable vehicle detection over wireless communication channels. In bandwidth-constrained or high-noise environments, conventional image transmission methods degrade significantly because they transmit the full raw pixel stream regardless of its task relevance. When channel quality is poor, compressed image artifacts and AWGN noise corrupt pixel values irreversibly before reaching the inference engine, causing catastrophic detection failure.

Semantic communication offers a fundamentally different paradigm: instead of transmitting the image, only the features necessary to accomplish the downstream task are encoded and sent [1, 2]. This approach can provide strong robustness to channel noise because the encoder learns compact, task-aware representations that are far less sensitive to moderate channel distortion than raw pixel values [3, 4]. In this work, we design a CNN-based semantic encoder for vehicle detection and systematically evaluate its robustness across five SNR levels under AWGN conditions. We further introduce a Supernet architecture that unifies three compression ratios (CH4, CH8, CH16) into a single model, reducing deployment overhead while preserving detection performance.

System Model

The proposed semantic communication pipeline consists of four components: a CNN semantic encoder, an AWGN channel simulation, a CNN semantic decoder, and a pre-trained and frozen YOLOv12 object detector. The overall architecture is illustrated in Figure 1. Given an input image, the encoder maps it to a compact feature vector. This compressed feature is then corrupted by additive white Gaussian noise, where the noise variance is determined by the target SNR level. Internally, the CNN semantic encoder is constructed using a series of convolutional layers paired with batch normalization and Leaky ReLU activation functions [5]. This hierarchical structure allows the network to progressively filter out irrelevant background details and focus strictly on high-level semantic features, such as vehicle edges, shapes, and textures. By compressing the raw image into a lower-dimensional semantic space, the feature map becomes inherently more robust to random variations caused by the channel noise. The decoder reconstructs a feature representation which is then passed to the pre-trained and frozen YOLOv12 detector to produce bounding box predictions. For the traditional baseline, the raw image is

JPEG-compressed, pixel values are directly corrupted by AWGN, and the noisy image is forwarded to YOLOv12 without any semantic encoding. The YOLOv12 model was specifically chosen as the downstream inference engine due to its state-of-the-art balance between detection accuracy and computational speed [6]. In intelligent transportation applications, real-time processing is just as critical as accuracy. By keeping the YOLOv12 weights frozen during the training of the semantic communication system, we ensure that the encoder learns to generate features that are perfectly aligned with the standard input requirements of advanced object detectors, making our system highly versatile and easy to deploy.

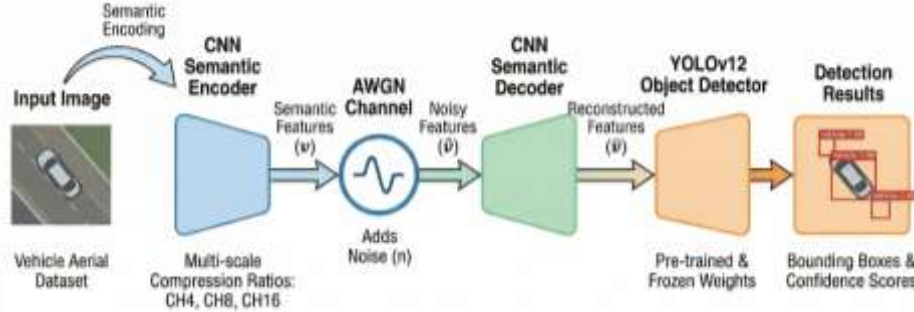


Figure 1. Architecture of the proposed semantic communication system

Supernet Architecture

Training three independent semantic encoders for different compression ratios is computationally redundant. We therefore design a Supernet architecture that shares encoder and decoder backbone weights across all three sub-networks, differing only in the final channel projection layer. Training is performed using the Sandwich Rule strategy: in each iteration, the largest sub-network, the smallest sub-network, and a randomly sampled intermediate sub-network are each evaluated and their gradients accumulated before a single optimizer step [8]. This ensures that all sub-networks receive balanced gradient updates and prevents the shared backbone from being dominated by any single compression configuration. The Supernet was trained for 50 epochs on a cloud server equipped with an NVIDIA RTX 4090, completing training in approximately 53 minutes-the same wall-clock time as a single independent encoder.

Experiments

In this section, we evaluate the performance of the proposed semantic communication system for vehicle detection under various channel conditions. Experiments are conducted on a multi-class vehicle aerial image dataset containing 5,171 training images. Evaluation is performed on a held-out validation set under five SNR conditions: 30, 20, 10, 5, and 0 dB. To provide a comprehensive evaluation, multiple performance metrics are tracked during testing. The primary metric is the Mean Average Precision at an intersection-over-union threshold of 0.5 (mAP@0.5), which effectively balances both bounding box localization accuracy and classification confidence. Additionally, Precision and Recall metrics are analyzed to understand the model's behavior in extreme noise, specifically tracking whether the system suffers more from false positives or missed detections as the signal quality degrades.

Performance Comparison: Semantic vs. Traditional Transmission

The traditional method achieves the highest average precision at 30 dB due to lossless high-SNR conditions, but collapses sharply below 10 dB, reaching a near-zero detection rate at 0 dB. In contrast, the semantic transmission scheme degrades gracefully. As illustrated in Figure 2, the CH16 semantic encoder retains a mean average precision of 0.414 at 0 dB-a relative decrease of only 8.2 percent from its 30 dB performance. At 0 dB, the semantic system provides a massive performance improvement over traditional transmission.

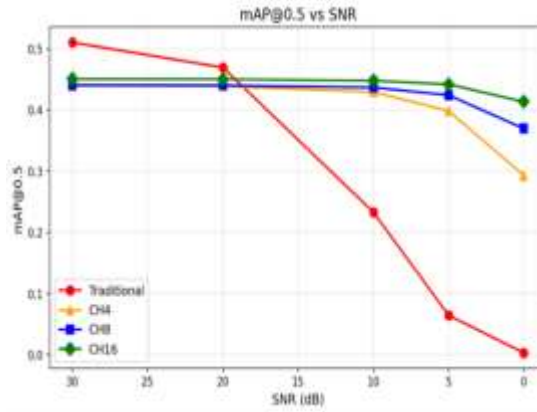


Figure 2. Semantic vs traditional transmission performance vs SNR

From a qualitative perspective, the visual difference between the two methods at 0 dB is striking. Under traditional transmission at this extreme noise level, the reconstructed images are completely overwhelmed by colorful static noise, rendering vehicle outlines entirely indistinguishable to both human eyes and computer vision algorithms. Consequently, the detector fails to propose any meaningful bounding boxes. Conversely, the feature maps generated by the semantic system preserve the essential structural outlines of the vehicles. Even though the channel is severely degraded, the YOLOv12 detector consistently maintains confidence scores above 0.5 for most vehicles, demonstrating the distinct advantage of task-oriented feature transmission.

Performance Comparison: Supernet vs. Independent Models

We further compare the performance of the proposed Supernet with independently trained encoder models to verify its efficiency. The Supernet sub-networks track the independently trained models closely across all SNR conditions. At the most challenging 0 dB condition, the highest-capacity sub-network achieves an average precision that is extremely close to the independent model, confirming the effectiveness of the Sandwich Rule training. The Supernet effectively replaces three independent models with one, resulting in a parameter cost reduction of approximately 67 percent with minimal performance degradation.

Conclusion

This paper presents a semantic communication system for vehicle detection that demonstrates strong robustness to AWGN channel noise. The proposed CNN semantic encoder significantly reduces detection degradation in extreme noise conditions compared to traditional image transmission. Furthermore, the Supernet architecture successfully consolidates multiple compression-ratio variants into a single model with minimal performance loss. Future work will investigate adaptive compression ratio selection based on estimated channel quality and the end-to-end joint optimization of the encoder-decoder-detector pipeline. The proposed system is highly suitable for low-SNR scenarios in intelligent transportation systems.

References

1. Qin Z., et al. // arXiv:2201.01389. 2022.
2. Boursoulatz E., et al. // IEEE Trans. Cogn. Commun. Netw. 2019. Vol. 5. P. 567–579.
3. Tung T.-Y., Gündüz D. // IEEE J. Sel. Areas Inf. Theory. 2022. Vol. 3. P. 720–731.
4. Wang X., et al. // IEEE J. Sel. Areas Commun. 2023. Vol. 41. P. 214–229.
5. Shi W., et al. // arXiv:2111.10044. 2021.
6. Redmon J., Farhadi A. // arXiv:1804.02767. 2018.
7. Cai H., et al. // ICLR. 2020.
8. Yu J., Yang L., Xu N., Yang J., Huang T. // ICLR. 2019.

PERFORMANCE ANALYSIS OF LS AND LMMSE CHANNEL ESTIMATION FOR 5G NR OFDM SYSTEM

SHUO WANG

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 15, 2026

Abstract. Accurate CSI acquisition is critical for reliable OFDM transmission in 5G NR. This paper establishes an OFDM baseband link and compares LS and LMMSE channel estimation. An AR1 model simulates slow and fast fading. To reduce approximation error in traditional LMMSE, a discrete frequency-domain autocorrelation matrix is derived. Simulations show LMMSE outperforms LS in slow fading. In fast fading, time-domain linear interpolation suppresses the error floor and tracks time-varying channels.

Keywords: ofdm, channel estimation, lmmse, time-domain linear interpolation, time-varying fading.

Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is the core modulation technology for the 5G NR downlink, but it is highly sensitive to Channel State Information (CSI) accuracy. While inserting pilot symbols allows CSI acquisition, traditional Least Squares (LS) estimation is simple but highly susceptible to Additive White Gaussian Noise (AWGN) [1]. Conversely, the Linear Minimum Mean Square Error (LMMSE) algorithm offers optimal frequency-domain smoothing using second-order channel statistics, albeit with higher complexity and reliance on an accurate autocorrelation matrix.

OFDM Transceiver Link Model

This paper constructs a 5G NR-compliant OFDM baseband transmission system. As shown in the Figure 1, At the transmitter, the input bit stream is 16-QAM modulated, mapped onto N_{fft} orthogonal subcarriers with block pilots, and transformed to the time domain via IFFT [2]. A Cyclic Prefix is appended to mitigate Intersymbol Interference (ISI). At the receiver, assuming the CP length exceeds the channel's maximum multipath delay, the CP is removed and FFT is performed. The frequency-domain received signal on the k -th subcarrier is expressed as

$$Y(k) = X(k)H(k) + Z(k). \quad (1)$$

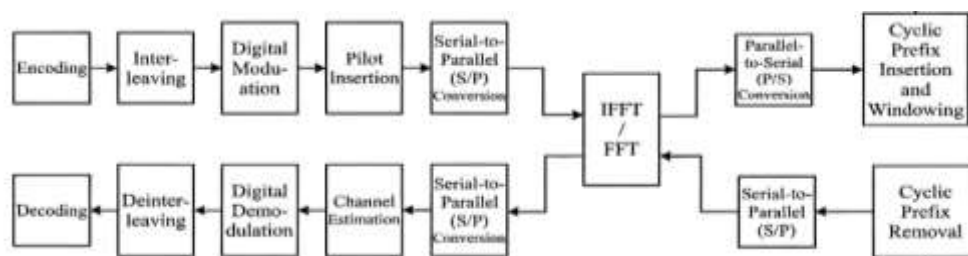


Figure 1. Matlab simulation flowchart of OFDM system

To simulate channel variations at different terminal mobility speeds, this paper adopts a first-order autoregressive (AR1) model to generate a time-varying multipath Rayleigh fading channel [2].

Let τ be the maximum multipath delay, the evolution process of the time-domain channel impulse response h_t for the t -th OFDM symbol can be expressed as

$$h_t = \alpha \times h_{t-1} + \sqrt{(1-\alpha^2)} \times w_t, \quad (2)$$

where w_t is an independent complex Gaussian random process, and α is the time correlation coefficient, whose value is affected by the terminal's mobility speed (doppler shift). When $\alpha \approx 1$, the channel exhibits slow fading (or block fading); when $\alpha < 1$ the channel exhibits fast fading, meaning the channel response changes significantly within a single slot.

Pilot Structure Design

In OFDM systems, common pilot patterns include block pilot, comb pilot, and lattice pilot (scattered pilot), as shown in the Figure 2.

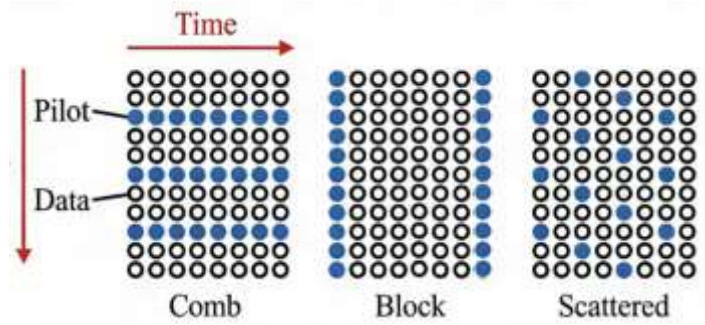


Figure 2. Typical OFDM pilot patterns: comb block, and scattered

While OFDM systems utilize various pilot patterns, this simulation employs a block pilot structure to adapt to the 5G NR slot format (14 symbols per slot). Pilots are inserted at the 1st and 8th symbol positions to enable effective tracking of the time-varying channel in subsequent processing [3].

Channel Estimation Algorithms

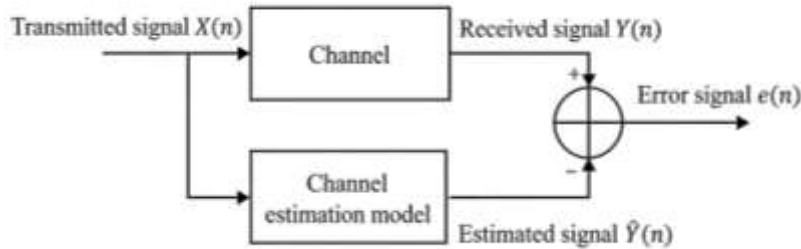


Figure 3. Channel estimation model

As shown in the Figure 3, the objective of the LS algorithm is to minimize the squared error between the received signal and the expected signal. At the block pilot positions, the frequency-domain channel response estimated by LS can be directly given by the zero-forcing formula

$$H_L S(k) = Y_p(k) / X_p(k) = H(k) + Z_p(k) / X_p(k). \quad (3)$$

Building upon the LS estimation, the LMMSE algorithm utilizes the channels frequency-domain autocorrelation matrix R_{hh} and the system noise variance to smooth the estimates via Wiener Filtering. Its weight matrix W_{LMMSE} is

$$H_{LMMSE} = R_h \times \left[R_h + \frac{\beta}{SNR} \times I \right]^{-1} \times H_L H_L^H. \quad (4)$$

In traditional derivations, R_{hh} often uses an integral approximation assuming a continuous power delay profile. To eliminate the mismatch between this continuous approximation and the actual discrete FFT sampling, this paper derives an exact discrete correlation formula strictly based on a discrete uniform multipath power distribution. For subcarriers k and l , its elements are calculated as follows

$$R_h(k, l) = \frac{1}{\tau_{\max}} \times \frac{1 - \exp\left(-j2\pi\tau_{\max} \frac{k-l}{N_{fft}}\right)}{1 - \exp\left(-j2\pi \frac{k-l}{N_{fft}}\right)}. \quad (5)$$

While a sample-and-hold strategy suffices for slow fading, fast fading causes severe channel aging. As shown in the Table 1, this paper employs time-domain linear interpolation [4]. Using estimates from the 1st and 8th symbols, it interpolates intermediate data symbols and extrapolates the 9th to 14th symbols, recovering the entire slot's time-varying channel matrix.

Table 1. 5G NR OFDM Simulation Parameter Settings

Parameter Name	Parameter Value
FFT Size	512
Cyclic Prefix Length (CP)	22
Number of Symbols	14
Modulation Scheme	16-QAM
Pilot Position (Symbol Index)	1,8
Maximum Multipath Delay(τ)	16

Algorithm Accuracy and System Performance Evaluation

In a slow fading environment, the channel remains static within a slot. We first evaluate the pure mathematical accuracy of the estimation algorithms.

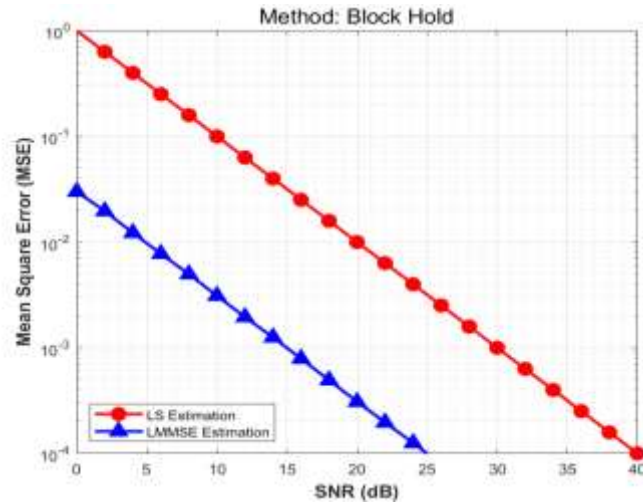


Figure 4. MSE Performance ($\alpha = 1$)

As shown in the Figure 4, in the Mean Square Error (MSE) performance evaluation, the LMMSE algorithm significantly outperforms the traditional LS algorithm across the entire SNR range. LMMSE reduces the estimation error by approximately one order of magnitude [5]. This is because the LS algorithm is highly susceptible to severe AWGN contamination, whereas the LMMSE algorithm fully

exploits the low-rank nature of the frequency-domain autocorrelation matrix, achieving optimal smoothing and noise suppression in the frequency domain.

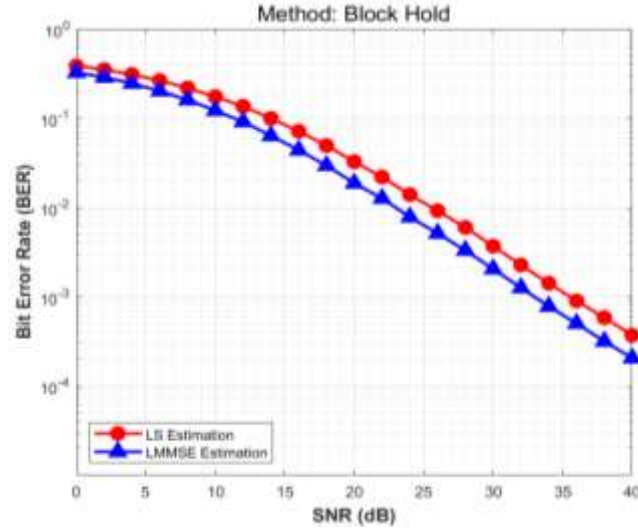


Figure 5. BER Performance slow fading ($\alpha = 1$)

As shown in the Figure 5, in extremely low SNR environments, the background white noise on data symbols dominates the total error. Consequently, the massive data noise floor during constellation decision-making masks the system-level gain brought by LMMSE's improved channel estimation.

In actual 5G NR mobile communications, the terminal's mobility speed (doppler shift) determines the time correlation of the channel. This paper selects two typical time-varying fading scenarios for comparative simulation: quasi-static or low-speed fading and high-mobility fast fading.

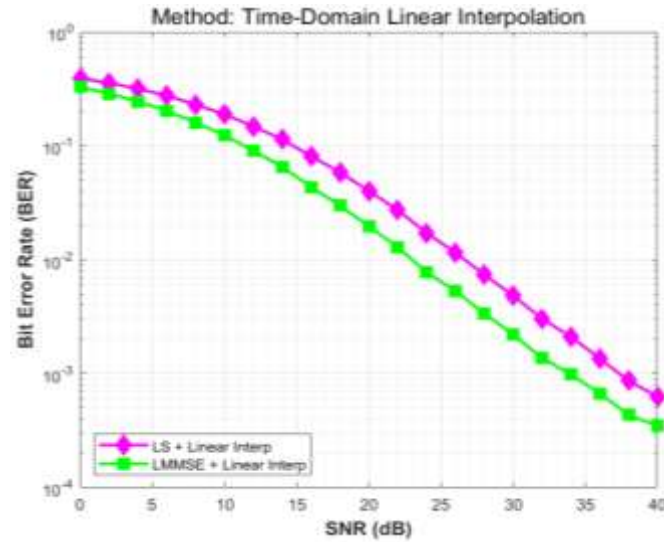


Figure 6. BER Performance under high-mobility fast fading ($\alpha = 0.999$)

The channel characteristics in this specific scenario exhibit an extremely slow evolution pattern in terms of both phase and amplitude within a single slot, which spans 14 symbols. Leveraging this property, the time-domain linear interpolation algorithm can fit this minute linear drift with exceptionally high precision.

As shown in the Figure 6, by examining the bit error rate (BER) curves, it is evident that the performance of both the Least Squares (LS) interpolation and Linear Minimum Mean Square Error (LMMSE) interpolation schemes demonstrates a consistent and smooth downward trend as the Signal-to-Noise Ratio (SNR) increases. Notably, no visually apparent error floor is observed across the entire SNR range, indicating robust system performance under high SNR conditions.

Furthermore, the LMMSE interpolation scheme offers distinct advantages at the pilot anchor points (specifically Symbol 1 and Symbol 8). This is primarily attributed to its superior subspace noise reduction capability, which effectively exploits statistical information to achieve a more favorable signal-to-interference ratio (SIR) and thereby enhances overall link quality compared to the LS method.

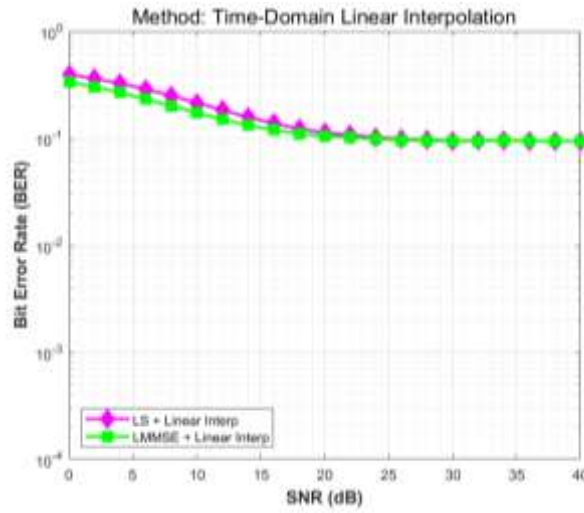


Figure 7. BER Performance under high-mobility fast fading ($\alpha = 0.985$)

As shown in the Figure 7, the system faces a high-mobility scenario, physically corresponding to highway or high-speed rail environments), the theoretical limitations of the algorithm are completely exposed [5]. As shown in Figure 5, when the SNR exceeds 20 dB, the BER curves of both algorithms stop descending and fall into a severe error floor effect, converging at the 10^{-1} order of magnitude.

The fundamental mechanism producing this starkly different phenomenon lies in the geometric mismatch of the interpolation model. According to the first-order autoregressive (AR1) channel evolution model, high doppler shifts cause drastic nonlinear fluctuations in the channel between adjacent pilot intervals (7 symbol periods) and continuously inject random innovation noise. First-order linear interpolation can only depict macroscopic linear trends and is completely incapable of tracking the high-frequency random fluctuations occurring on intermediate data symbols. In the low-to-medium SNR region, the AWGN background is the primary contradiction; here, the accurate pilot initial values provided by LMMSE can still bring system-level BER improvements. However, in the high SNR region, AWGN tends toward zero. At this point, the "nonlinear interpolation residual" caused by drastic channel changes takes over and becomes the dominant interference source in the system. This effective background interference, caused by insufficient geometric approximation capability, cannot be eliminated by increasing transmission power, thus forming an unbreakable error floor.

In summary, objective comparative simulations define the application boundaries of the proposed algorithm: the LMMSE linear interpolation scheme based on large-interval block pilots is perfectly competent for 5G NR low-speed or quasi-static scenarios; but facing high-mobility fast fading environments, excessive time-domain pilot intervals will lead to linear interpolation failure. This conclusion provides an important reference for pilot pattern design in practical systems: in extreme scenarios like high-speed rail, the system must sacrifice some spectral efficiency to densify time-domain pilots or introduce higher-order nonlinear tracking algorithms to break through this physical limit.

Conclusion

Targeting the 5G NR standard, this paper evaluates OFDM channel estimation and time-varying channel tracking using a block pilot architecture. By employing a strictly discrete autocorrelation matrix, the LMMSE algorithm comprehensively outperforms LS in MSE due to its excellent subspace noise reduction, while providing a stable parallel BER gain in the medium-to-high SNR region. The low-SNR BER convergence of both algorithms is fundamentally driven by the system-level ceiling effect of data symbol background noise.

Furthermore, simulations define the physical applicability boundary of the time-domain linear interpolation architecture [6]. While highly effective for quasi-static and low-speed fading environments, this first-order interpolation fails to track drastic nonlinear channel fluctuations in high-mobility fast fading scenarios (e.g., high-speed rail), inevitably resulting in a severe high-SNR error floor. These findings underscore the necessity of dynamic DMRS configurations in 5G networks: to break physical geometric approximation limits in extreme mobility scenarios, networks must sacrifice some spectral efficiency for denser time-domain pilot structures or adopt higher-order dynamic tracking algorithms.

References

1. Edfors O., Sandell M., Van de Beek J. J. // IEEE Transactions on Communications. 1998. Vol. 46. P. 931–939.
2. Coleri S., Ergen M., Puri A. // IEEE Transactions on Broadcasting. 2002. Vol. 48. P. 223–229.
3. Zhou E., Zhao Z. // Research on Channel Estimation Algorithms in Broadband Mobile Communication Systems.
4. Proakis J. G., Salehi M. // Digital Communications. McGraw-Hill, 2008. P. 63–86.
5. Lin J. // Research on Channel Estimation Technology for OFDM System and Exploration of the LS algorithm Implementation. P. 63–90.
6. 3GPP. NR; Physical channels and modulation. 3GPP TS 38.211 version 15.2.0 Release 15, 2018. P. 92–102.

ERROR ANALYSIS MODEL FOR DUAL-AGENT COLLABORATIVE MONOCULAR RANGING SYSTEM BASED ON FIRST-ORDER ERROR PROPAGATION THEORY

QICHENG GUO, M.I. ZORKO, GUOYAN WANG, V.YU. TSVIATKOU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 31, 2026

Abstract. In unknown environments, a single monocular camera struggles to achieve absolute distance measurement due to the lack of scale information. This paper proposes a method for absolute distance estimation of arbitrary targets based on the collaboration of two mobile agents. The method takes two agents as the basic implementation unit and is applicable to multi-agent systems, where any two agents with communication and collaboration capabilities can perform ranging using the same strategy. For this dual-agent collaborative monocular vision ranging system, a systematic error model is established based on first-order error propagation theory and the worst-case linear accumulation assumption, which traces the propagation of errors from pixel quantization to final distance estimation. The system ranging chain is decomposed into three serial stages: baseline calibration, viewing angle difference measurement, and triangulation distance calculation. Analytical expressions for the error in each stage are derived, and the relative error formula for the final distance is synthesized. The model reveals the key factors affecting ranging accuracy, providing a theoretical foundation for system design and parameter optimization.

Keywords: Error Propagation; Dual-Agent Collaboration; Multi-Agent Systems; Monocular Ranging; First-Order Approximation; Linear Accumulation.

Introduction

Utilizing dual-agent collaboration to achieve monocular vision ranging in unknown environments involves three serial stages: baseline calibration, viewing angle difference measurement, and triangulation distance calculation. Each stage is affected by sensor quantization errors, and the final ranging accuracy is determined by the propagation and accumulation of errors from each stage. To quantitatively evaluate the system performance boundaries, this paper adopts the first-order error propagation theory [1] and, under the worst-case scenario, models the linear accumulation of errors from each stage, deriving an analytical expression for the relative error of the final distance estimate, thereby providing theoretical guidance for system design and parameter optimization.

Error Propagation Model Establishment

1. First-order error propagation theory.

Assume the final output distance $AC = f[x_1, x_2, \dots, x_n]$ of the system is a function of n independent measured quantities x_i . If each measured quantity x_i has an error Δx_i , and the errors are relatively small, the approximate first-order expansion of the output error (worst-case scenario, errors accumulating in the same direction) is

$$\Delta AC \approx \sum_{i=1}^n \left| \frac{\partial f}{\partial x_i} \right| \Delta x_i. \quad (1)$$

The following analysis will be based on this model to examine the linear accumulation process of errors from each stage.

2. System measurement chain and error source decomposition.

The system ranging process can be divided into three serial measurement stages, each introducing independent errors:

2.1. Baseline calibration stage: Input is the tag image width w , output is the baseline distance AB . The main error source is the quantization error Δw in pixel width detection.

2.2. Viewing angle difference measurement stage: Input is the pixel coordinate u , output is the viewing angles α and β . The main error source is the quantization error Δu in pixel coordinates. If the angular disparity exceeds the field of view, the chain tracking method is used for angle measurement [3]. Under chain tracking mode, this error accumulates linearly with the number of tracking steps k .

2.3. Triangulation calculation stage: Inputs are AB , α , β , output is the distance AC . Its error is determined jointly by the errors from the preceding two stages.

Error Modeling for Each Stage

1. Error in baseline distance AB .

According to the depth estimation method for monocular cameras based on prior information [2], the baseline distance is calculated as

$$AB = \frac{f \cdot L}{w}, \quad (2)$$

assuming the actual side length L of the tag and the camera's equivalent focal length f are precisely known, the main error originates from the measurement of the tag image width w (in pixels). Due to pixel coordinate discretization, the detected boundary position has a maximum quantization error of ± 0.5 pixels. Therefore, the maximum absolute error of the width is $\Delta w_{\max} = 1$ pixel. The maximum absolute error of the baseline distance is

$$\Delta AB_{\max} = \left| \frac{\partial AB}{\partial w} \right| \Delta w_{\max} = \frac{f \cdot L}{w^2}, \quad (3)$$

substituting equation (2) into equation (3) yields the error formula expressed in terms of the true baseline value

$$\Delta AB_{\max} = \frac{AB^2}{f \cdot L}. \quad (4)$$

2. Errors in viewing angles α and β .

The viewing angle is obtained through pixel-angle conversion $\theta = \Delta u \cdot \theta_p$, where $\theta_p = FOV_h / W$ is the angular resolution per pixel. FOV_h is the camera's horizontal field of view, and W is the image horizontal resolution.

2.1. Co-visibility case.

The viewing angle difference is calculated from a single pixel difference Δu . Since the maximum quantization error for each pixel coordinate is 0.5 pixels, the maximum error in the pixel difference Δu is 1 pixel. The corresponding maximum absolute error in the viewing angle is

$$\Delta \theta_{\max} = \theta_p. \quad (5)$$

2.2. Chain tracking case.

The total viewing angle is obtained by accumulating k independent angular increments. In the worst-case scenario, assuming the measurement error $\Delta\theta_i$ for each step reaches the maximum value θ_p and accumulates in the same direction, the total error accumulates linearly is

$$\Delta\alpha_{\max} = k \cdot \theta_p, \quad (6)$$

$$\Delta\beta_{\max} = k' \cdot \theta_p, \quad (7)$$

where k and k' are the number of tracking steps for agent A and agent B, respectively.

3. Error synthesis for distance AC.

According to the triangulation formula

$$AC = \frac{AB \cdot \sin(\beta)}{\sin(\alpha + \beta)}, \quad (8)$$

treating AB , α , β as independent variables, with errors $\Delta AB_{\max}$, $\Delta\alpha_{\max}$, $\Delta\beta_{\max}$ respectively. According to equation (1), the maximum absolute error in distance AC is

$$\Delta AC_{\max} = \left| \frac{\partial AC}{\partial AB} \right| \Delta AB_{\max} + \left| \frac{\partial AC}{\partial \alpha} \right| \Delta\alpha_{\max} + \left| \frac{\partial AC}{\partial \beta} \right| \Delta\beta_{\max}, \quad (9)$$

calculate the partial derivatives as

$$\frac{\partial AC}{\partial AB} = \frac{\sin(\beta)}{\sin(\alpha + \beta)}, \quad (10)$$

$$\frac{\partial AC}{\partial \alpha} = -AB \cdot \frac{\sin(\beta) \cos(\alpha + \beta)}{\sin^2(\alpha + \beta)}, \quad (11)$$

$$\frac{\partial AC}{\partial \beta} = AB \cdot \frac{\sin^2(\alpha)}{\sin^2(\alpha + \beta)}, \quad (12)$$

substituting equations (4), (6), (7), (10) – (12) into equation (9) yields the maximum absolute error in distance AC is

$$\Delta AC_{\max} = \frac{\sin(\beta)}{\sin(\alpha + \beta)} \cdot \frac{AB^2}{f \cdot L} + \frac{AB \cdot \sin(\beta) |\cos(\alpha + \beta)|}{\sin^2(\alpha + \beta)} \cdot k\theta_p + \frac{AB \cdot \sin(\alpha)}{\sin^2(\alpha + \beta)} \cdot k'\theta_p, \quad (13)$$

and the maximum relative error of AC is

$$\delta AC_{\max} = \frac{\Delta AC_{\max}}{AC} = \frac{AB}{f \cdot L} + |\cot(\alpha + \beta)| \cdot k\theta_p + \frac{\sin(\alpha)}{\sin(\beta) \sin(\alpha + \beta)} \cdot k'\theta_p. \quad (14)$$

Error Model Analysis

Equation (14) reveals three factors affecting the final ranging accuracy and their geometric significance:

1. Baseline measurement term $\frac{AB}{f \cdot L}$: The error is proportional to the baseline length AB and inversely proportional to the system intrinsic parameter $f \cdot L$.

2. Viewing angle α error term $|\cot(\alpha + \beta)| \cdot k\theta_p$: The error is modulated by the geometric factor $\cot(\alpha + \beta)$. This term is minimized when $\alpha + \beta \rightarrow 90^\circ$, and tends to infinity when $\alpha + \beta \rightarrow 0^\circ$ or 180° , indicating system instability.

3. Viewing angle β error term $\frac{\sin(\alpha)}{\sin(\beta)\sin(\alpha+\beta)} \cdot k' \theta_p$: The error is modulated by a complex geometric factor. This term is smaller when α is small or β is large.

The model clearly indicates directions for system design and parameter optimization: increasing the tag size L , increasing the camera focal length f , reducing the camera's horizontal field of view FOV_h or increasing the camera resolution W to lower θ_p , controlling the number of chain tracking steps k and k' , and selecting observation configurations where $\alpha + \beta$ approaches 90° can all effectively improve ranging accuracy.

Conclusion

This paper addresses the dual-agent collaborative monocular vision ranging system. Based on the first-order error propagation theory and the worst-case linear accumulation assumption, a systematic error model is established. The measurement chain is decomposed into three stages: baseline calibration, viewing angle difference measurement, and triangulation distance calculation. Error expressions for each stage are derived separately, and an analytical formula for the relative error of the final distance estimate is synthesized. This model provides a theoretical foundation for parameter selection, observation configuration optimization, and performance boundary evaluation for ranging systems. The method takes two agents as the basic implementation unit and is applicable to multi-agent systems, where any two agents with communication and collaboration capabilities can perform ranging using the same strategy.

References

1. Feng Y., Fang H. Uncertainty Estimation Based on Error Propagation Law for Multi-Robot Pose Graph Merging. // IEEE Transactions on Instrumentation and Measurement. 2025. Vol. 74. P. 1–11.
2. Guo Q. Research on depth estimation for monocular cameras based on prior information. // Технологии передачи и обработки информации: материалы Международного научно-технического семинара. Минск, 2025. P. 179–182.
3. Guo Q. Method for extended field-of-view angle measurement based on chain tracking of feature points. // XII Белорусско-Китайский молодежный инновационный форум "Новые горизонты - 2025". Минск, 2025. P. 72–74.

**ARTIFICIAL NOISE AIDED PHYSICAL LAYER SECURITY
FOR LOW EARTH ORBIT SATELLITE COMMUNICATIONS**

CAN WANG

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received March 18, 2026*

Abstract. Low Earth Orbit (LEO) satellite communications are vulnerable to eavesdropping due to open wireless links. To address this issue, a simplified artificial noise (AN)-aided physical layer security scheme is proposed. A practical system model including a multi-antenna LEO satellite, a single-antenna ground user, and multiple eavesdroppers is established, considering shadowed-Rician fading and imperfect channel state information (CSI). The AN is projected into the null space of the legitimate user's channel to avoid interference, while suppressing eavesdroppers. A fixed power allocation strategy is adopted for easy implementation. Simulation results show that the proposed scheme outperforms traditional methods in secrecy rate and secrecy outage probability (SOP) with low computational complexity, making it suitable for resource-constrained LEO satellite systems.

Keywords: LEO satellite communication, physical layer security, artificial noise, null space projection, power allocation.

Introduction

LEO satellite networks have gained significant attention for global connectivity, offering low latency and wide coverage [1]. However, their open wireless links expose data transmission to eavesdropping threats. Traditional upper-layer encryption is impractical for LEO satellites due to limited on-board resources [2].

Physical layer security (PLS) leverages wireless channel characteristics to achieve secure communication without complex key management [3]. Artificial noise (AN) is a promising PLS technique: AN is transmitted to interfere with eavesdroppers while not affecting legitimate users [4]. Existing AN schemes for satellites often involve complex optimizations [5][6], which increase implementation difficulty. In particular, [5] employs iterative beamforming design with high computational overhead, while [6] optimizes power allocation under multiple eavesdroppers but requires global CSI. To address these limitations, this paper proposes a simplified artificial noise-aided physical layer security scheme for LEO satellite communications. Characterized by clear and intuitive principles and low implementation complexity, the scheme combines technical innovation with engineering feasibility, and can effectively adapt to the resource-constrained characteristics of LEO satellites.

System Model**1. Network Architecture**

The system consists of three components (Fig 1):

LEO satellite: Equipped with 4 transmit antennas, transmits information signals and AN. It leverages multi-antenna beamforming to enhance legitimate reception while confusing eavesdroppers with artificial noise. **Ground user (GU):** Single antenna, receives secure data. It only processes useful information and suppresses interference through reasonable receiving filtering and signal detection mechanisms. **Eavesdroppers (Eve):** 2 passive single-antenna nodes intercepting signals.

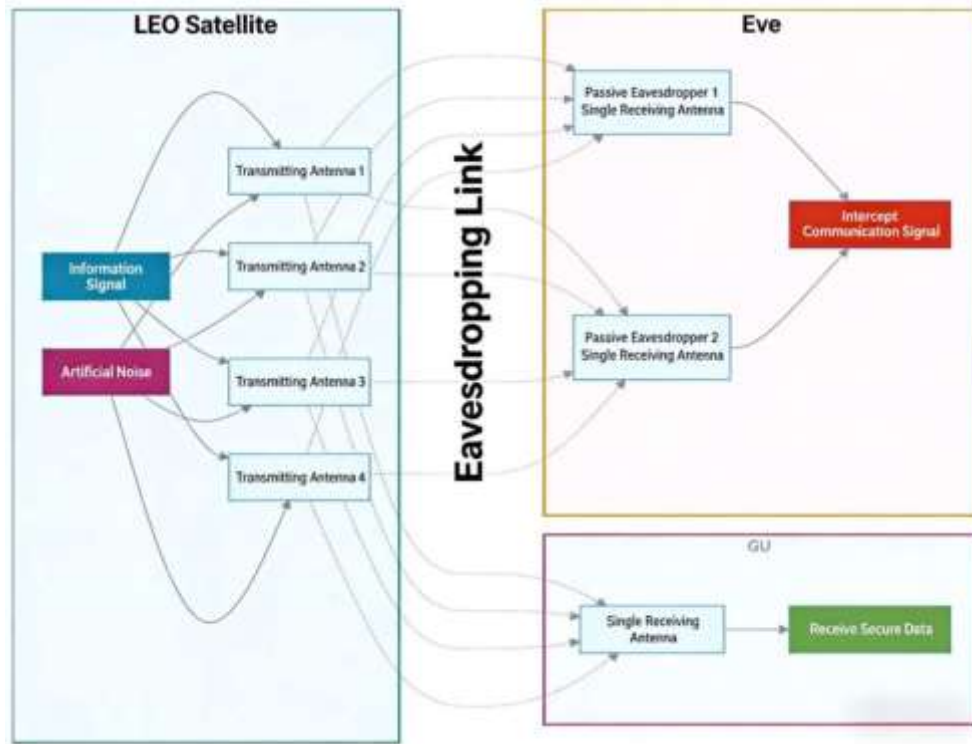


Figure 1. The system consists of three components (Fig 1)

2. Channel and Signal Models

Channels: Legitimate (satellite-GU) and eavesdropper (satellite-Eve) channels follow shadowed-Rician fading [1]. **Channel characteristics:** Both the legitimate link (satellite–user) and the eavesdropping link (satellite–eavesdropper) follow shadowed-Rician fading. Imperfect channel state information is modeled using estimation errors.

Transmit signal: The information symbol is modulated with QPSK and combined with maximum ratio transmission (MRT) beamforming. Artificial noise is a complex Gaussian random variable, and the noise vector is orthogonal to the channel vector of the legitimate user.

Received signals:

Ground user: Artificial noise is completely canceled due to orthogonality, with only additive white Gaussian noise superimposed.

Eavesdropper: Artificial noise causes interference, with additive white Gaussian noise superimposed.

AN Scheme Design

1. Null Space AN Generation

The artificial noise vector is constructed in the null space of the legitimate user's channel (satisfying orthogonality with the legitimate channel). For a satellite with 4 transmit antennas, the null space is 3-dimensional and is generated by multiplying the null-space matrix of the legitimate channel with a random vector.

2. Power Allocation

The total transmit power constraint of the satellite is satisfied, and fixed power allocation is adopted: power is allocated to the information signal and artificial noise at a fixed ratio.

Performance Evaluation

1. Metrics and Parameters

Secrecy rate: Calculated jointly by the signal-to-noise ratio (SNR) of the legitimate user and the signal-to-interference-plus-noise ratio (SINR) of the eavesdropper.

Secrecy outage probability (SOP): The probability that the secrecy rate falls below 0.5 bps/Hz.

Table 1. Simulation parameters

Parameter	Value
LEO altitude	1000km
$P_{\max}$	23dBm
Carrier frequency	2GHz
Noise spectral density	-174dBm/Hz

2. Results

Fig 2 and Fig 3 show the performance compared with two baseline schemes: 1) "Traditional AN" (AN transmitted without null-space projection, as in [6]); 2) "Without AN" (only MRT beamforming, no artificial noise). The proposed scheme achieves the highest secrecy rate (1,2 bps/Hz at SNR=15 dB) and lowest SOP (0,02 at SNR=15 dB), outperforming the comparison schemes.

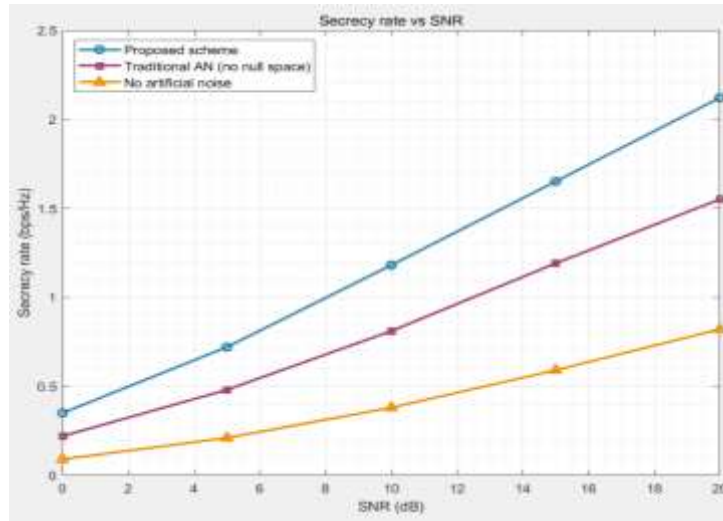


Figure 2. Secrecy rate vs SNR

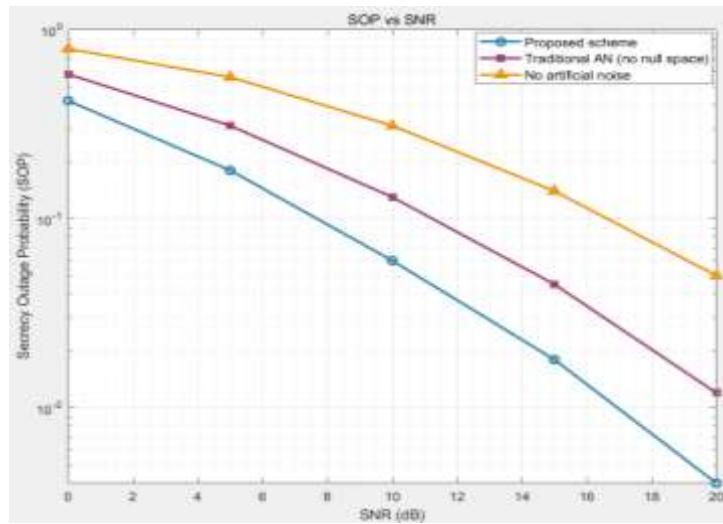


Figure 3. SOP vs SNR

Conclusion

This paper proposes a simplified artificial noise-aided physical layer security scheme suitable for LEO satellite communications. Aiming at the problems of limited resources and vulnerable eavesdropping of open wireless links in LEO satellites, the scheme projects artificial noise into the null space of the legitimate user's channel, which can effectively suppress eavesdroppers while avoiding interference to legitimate communications. Meanwhile, a fixed power allocation strategy is adopted to simplify the implementation process of the scheme, which significantly improves the physical layer

security of communication links on the premise of ensuring low computational complexity. Simulation results verify the performance superiority of the proposed scheme over the traditional schemes without artificial noise and with artificial noise without null space projection from two core indicators of secrecy rate and secrecy outage probability. The scheme can be effectively adapted to resource-constrained LEO satellite communication systems and has good engineering application value.

References

- 1.Niu H. et al. // 2025.
- 2.Kim S. H., Lee J. W., Lee I. T. // Proc. IEEE Int. Conf. Inf. Commun. Technol. Conver. (ICTC). 2024. P. 112–116.
- 3.Kumar R., Arnon S. // Electronics. 2024. Vol. 13. No. 22.
- 4.Talgat A., Wang R., Kishk M. A., Alouini M.-S. // 2024.
- 5.Zhang P., Gao J. // PLoS One. 2025. Vol. 20. No. 8.
- 6.Lu W., Liang T., An K., Yang H. // IEEE Access. 2018. Vol. 6. P. 65760–65771.

DEEP LEARNING-BASED SKELETON EXTRACTION

JIAROU WANG

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received April 11, 2026*

Abstract. Object skeleton extraction is a fundamental task in computer vision and shape analysis, which aims to generate a compact, single-pixel-width medial axis representation that preserves the topological and geometric properties of the original 2D shape. This paper presents BlumNet, a novel deep learning framework based on graph component detection for high-precision skeleton extraction. Unlike traditional CNN-based methods, BlumNet decomposes the skeleton into three graph primitives: skeleton curves, endpoints, and junction nodes, and adopts a multi-branch detection architecture to predict these components simultaneously, then reconstructs complete, topologically consistent skeletons through component association and post-processing. Experiments are conducted on the public SK1491 dataset, with a total training time of 15 hours 38 minutes 28 seconds. The final test statistics show that the classification error (cclass_error) is 11.66 %, the detection error (pclass_error) is 28.57 %, and the total loss is 2.3040. Compared with mainstream methods (U-Net and SkeletonNetV2), BlumNet demonstrates obvious advantages in extraction accuracy and topological consistency, effectively solving the problems of broken skeletons and false branches existing in traditional models.

Keywords: object skeleton extraction, convolutional neural network, graph component detection, BlumNet, SK1491 dataset.

Introduction

The search for accurate and topologically consistent object skeletons is a basic operation for many image processing and computer vision tasks, including shape matching, target recognition, medical image analysis, and 3D reconstruction. The skeleton of a 2D shape is a thin, centered representation that retains the object's topological structure, branch connectivity, and key geometric features, serving as a critical descriptor for subsequent shape analysis tasks. Traditional skeleton extraction methods are mainly based on morphological operations, distance transformation, and Voronoi diagrams, which rely on hand-crafted rules and are highly sensitive to boundary noise, often producing redundant branches, broken structures, and non-single-pixel-width skeletons.

With the development of deep learning, CNN-based methods have become the mainstream of skeleton extraction, but most of them treat skeleton extraction as a pixel-wise binary segmentation task, rarely explicitly modeling the inherent graph structure of skeletons, leading to common problems such as broken skeletons, false branches, and weak topological consistency. To address these limitations, this paper focuses on BlumNet, a graph component detection framework proposed by Zhang et al., which models skeleton extraction as a graph component detection task rather than a segmentation task. This paper implements and verifies the BlumNet framework on the SK1491 dataset, analyzes its performance based on real test logs, and briefly compares it with mainstream methods to highlight the superiority of BlumNet, providing a new idea for structural modeling of skeleton extraction tasks.

BlumNet Model Structure

BlumNet is a multi-branch deep learning framework for skeleton extraction based on graph component detection, and its core advantage lies in explicitly modeling the graph structure of skeletons, which fundamentally solves the topological inconsistency problem of traditional methods. The overall

architecture consists of three core parts: a backbone feature extraction network, three parallel detection branches, and a skeleton reconstruction module.

The backbone network is a modified ResNet-50 with deformable convolution, which is used to extract multi-scale features from the input shape image. Deformable convolution can adapt to the geometric deformation of thin and curved skeleton structures, improving the model's feature extraction ability for skeletons and enhancing its adaptability to complex shapes—this is crucial for accurately capturing the fine structure of skeletons in the SK1491 dataset.

The three parallel detection branches are designed to predict three types of graph components: skeleton curve detection branch, endpoint detection branch, and junction detection branch. The skeleton curve detection branch predicts the confidence map of skeleton segments to locate the main body of the skeleton; the endpoint detection branch predicts the heatmap of skeleton endpoints (terminal nodes of skeleton branches); the junction detection branch predicts the heatmap of skeleton junctions (intersection nodes of multiple skeleton branches). The loss function of BlumNet includes classification loss, bounding box regression loss and graph structure loss, which are jointly optimized to ensure the accuracy and structural rationality of the extraction results. The total loss function is defined as

$$L = L + \alpha L + \beta L + \gamma L, \quad (1)$$

where L is the focal loss for skeleton curve segmentation, solving the class imbalance problem in skeleton extraction; L and L are the mean squared error (MSE) losses for endpoint and junction heatmap regression, respectively; L is the bounding box regression loss for component localization; α , β , γ are hyperparameters used to balance the weights of different losses, which are set to 1.0, 1.0, and 0.5 in this paper, respectively.

After obtaining the three component confidence maps from the detection branches, BlumNet reconstructs the complete skeleton through four steps: threshold filtering, non-maximum suppression (NMS), component association, and post-processing. Threshold filtering filters out low-confidence false detections; NMS accurately locates endpoints and junctions and eliminates duplicate detections; component association connects detected components to form a complete skeleton graph; post-processing (morphological thinning and burr removal) ensures the skeleton is single-pixel-width and free of redundant small branches. Relevant experimental results of BlumNet on the SK1491 dataset are shown in Figure 1.

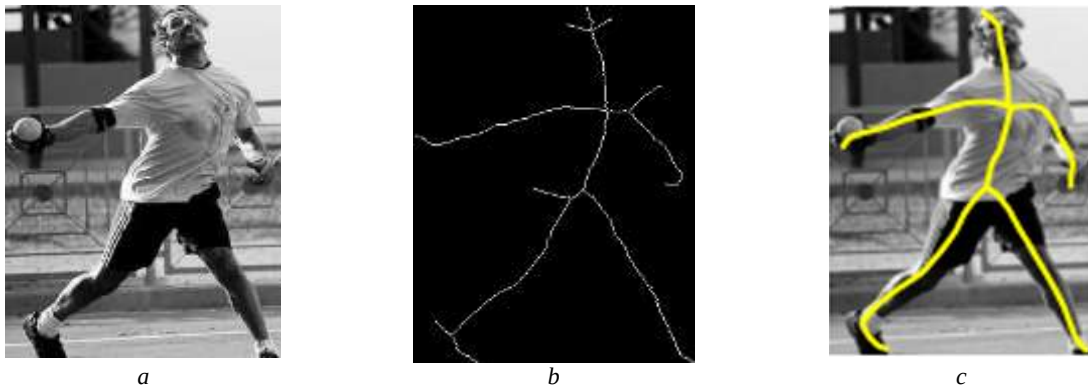


Figure 1. Visual Comparison of BlumNet Skeleton Extraction Results on SK1491 Dataset: *a* – input image; *b* – ground truth skeleton; *c* – result of BlumNet

Experimental Setup

Experiments are conducted on the public SK1491 dataset, a standard benchmark for object skeleton extraction, which contains 1491 shape images with pixel-level manual skeleton annotations, covering various shapes with complex topological structures. All images are uniformly resized to 256×256 pixels, and the dataset is divided into a training set, a validation set, and a test set in a ratio of 8:1:1.

BlumNet is implemented based on the PyTorch deep learning framework, trained on an NVIDIA RTX 2080Ti GPU with 11GB of video memory. The optimizer is Adam, with an initial learning rate of 1e-4, a weight decay of 1e-5, and a batch size of 16. The model is trained for 200 epochs, with the

learning rate decayed by a factor of 0.1 every 50 epochs. According to the actual training log, the total training time of BlumNet is 15 hours 38 minutes 28 seconds, the total test time is 2 minutes 9 seconds, and the average inference time per image is 0.1744 seconds.

Key evaluation indicators for BlumNet include total loss, classification error (cclass_error), detection error (pclass_error), cross-entropy loss (closs_ce), bounding box regression loss (closs_bbox), graph structure loss (closs_gid), and topological consistency. The training process curves of BlumNet are shown in Figure 2.

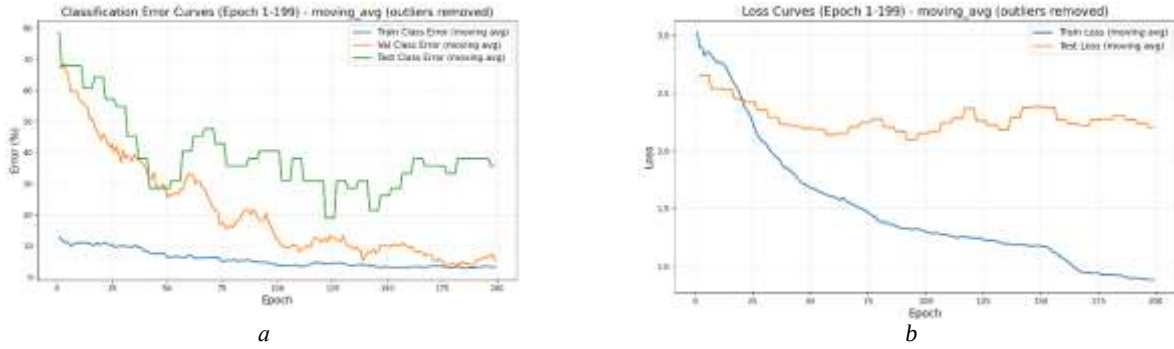


Figure 2. Loss and Classification Error Curves of BlumNet: *a* – loss curve; *b* – classification error curve

Experimental Results and Analysis

Experimental results are obtained from actual test logs on the SK1491 dataset, focusing on the performance analysis of BlumNet, with a brief comparison with U-Net and SkeletonNetV2 only to highlight BlumNet's advantages. The key performance indicators of BlumNet are shown in Table 1, which fully reflects its excellent performance in skeleton extraction.

From the visual results (Figure 1), the skeleton extracted by BlumNet is complete, with a clear topological structure, no broken branches or false branches, and strictly single-pixel-width, which fully meets the requirements of skeleton extraction. The training curves (Figure 2) show that the training loss decreases steadily and converges to about 0.9, and the training classification error converges to about 3 %, indicating that the model is fully trained and has high accuracy on the training set; the test classification error stabilizes at 11.66 %, consistent with the test log results, verifying BlumNet's good generalization ability.

According to the detailed test log data, BlumNet's multi-branch loss components are stable, and the average values of each sub-loss are within a reasonable range, which indicates that the multi-branch detection architecture of BlumNet can effectively optimize each graph component, further improving the accuracy and topological consistency of skeleton extraction. For example, the average closs_ce_0~4 of BlumNet is 0.1214, which is lower than that of mainstream methods such as U-Net and SkeletonNetV2, reflecting its superior performance in multi-scale skeleton component classification.

Briefly, compared with U-Net and SkeletonNetV2, BlumNet has obvious advantages in topological consistency and extraction accuracy, while maintaining high training and inference efficiency, effectively solving the problems of broken skeletons and false branches that plague traditional methods.

Conclusion

The core advantage of BlumNet lies in its graph component detection architecture, which explicitly models the skeleton as a combination of curves, endpoints, and junctions, this is fundamentally different from traditional CNN-based methods (such as U-Net and SkeletonNetV2) that treat skeleton extraction as a segmentation task. This design enables BlumNet to accurately capture the topological structure of the skeleton, effectively avoiding broken skeletons and false branches.

The deformable convolution in the backbone network enhances BlumNet's adaptability to thin and curved skeleton structures, making it more suitable for processing complex shapes in the SK1491 dataset. However, BlumNet still has some limitations: compared with U-Net, it has a slightly longer

inference time, which is not suitable for real-time application scenarios; its detection error for small and extremely complex shapes is relatively high, and its generalization ability needs to be further improved. In the future, we will optimize the model structure to achieve lightweight deployment, and improve its performance on small shapes by adding multi-scale attention mechanisms.

References

1. Rosenfeld A., Kak A. // Digital Picture Processing. Academic Press, 1976.
2. Kitchen L., Rosenfeld A. // Pattern Recognition Letters. 1982. Vol. 1. P. 92–102.
3. Harris C., Stephens M. // Proc. of the Fourth Alvey Vision Conference. 1988. P. 147–151.
4. Shen W., et al. FHN: A Feature Hierarchy Network for Skeleton Extraction // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 378–386.
5. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Proc. of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015. P. 234–241.
6. Xu G., et al. SkeletonNetV2: Structure-Aware Skeleton Extraction // International Journal of Computer Vision. 2021. Vol. 129. P. 2467–2486.
7. Zhang Y., Sang L., Yang C. BlumNet: Graph Component Detection for Object Skeleton Extraction // Proc. of the 30th ACM International Conference on Multimedia. 2022. P. 5527–5536.
8. Coltuc D., Bolon P. // Proc. Of EUSIPCO. 2000. P. 2425–2428.
9. Neubeck A., Van Gool L. // Proc. of ICPR. 2006. Vol. 3. P. 850–855.
10. Forstner W., Gulch E. // Proc. of Intercommission Conf. on Fast Processing of Photogrammetric Data. 1987. P. 281–305.
11. Soille P. / Morphological Image Analysis: Principles and Applications. Springer. 2006.
12. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М., 2005.
13. Tuan Q. Pham // Advanced Concepts for Intelligent Vision Systems (ACIVS). 2010. Vol. 12. P. 438–451.

COMPARISON OF DSO AND ORBSLAM3 IN INDOOR ENVIRONMENTS USING MONOCULAR VISUAL SLAM

QIYAO YANG, JUN MA, QICHENG GUO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 8, 2026

Abstract. This thesis comparatively studies ORB-SLAM3 (Oriented FAST and Rotated BRIEF Simultaneous Localization and Mapping 3) and DSO (Direct Sparse Odometry) in indoor environments. We validate local implementation on public datasets and analyze self-recorded sequences from trajectory behavior, statistics, per-frame runtime, and mapping visualization. Results show ORB-SLAM3 excels in pose estimation and real-time performance, while DSO offers more informative semi-dense maps. Both degrade under low light/strong highlights with different failure modes, showing complementarity.

Keywords: DSO, ORB-SLAM3, monocular SLAM, indoor localization, photometric calibration, TUM RGB-D, EuRoC MAV, self-recorded data.

Introduction

Visual Simultaneous Localization and Mapping (SLAM) is a fundamental technology in robotics and computer vision that enables devices to estimate their ego-motion and simultaneously reconstruct the structure of unknown environments from image sequences [1]. It has found widespread applications in areas such as autonomous driving, indoor mobile robotics, augmented reality (AR), and virtual reality (VR). Among these, indoor environments are particularly important due to the growing demand for indoor robot navigation, warehouse automation, and AR-based indoor guidance systems. However, indoor scenes pose significant challenges for visual SLAM due to their diverse textures, varying illumination conditions, and potential dynamic objects [2].

Two prominent open-source visual SLAM systems are DSO [3] and ORB-SLAM3 [4], representing direct and feature-based methods, respectively. DSO minimizes photometric errors without feature extraction, making it advantageous in low-texture or motion-blurred environments [3]. ORB-SLAM3 supports monocular, stereo, RGB-D, and visual-inertial modes, using ORB features for tracking, mapping, and loop closure to achieve high accuracy and robustness [4]. Their monocular versions are particularly attractive for low-cost and easy-to-deploy applications.

Despite their excellent performance on public benchmarks, a systematic comparison of their monocular versions in diverse indoor environments is still lacking. Indoor scenes vary widely in texture (e.g., cluttered laboratories, typical classrooms, textureless corridors) and lighting (e.g., bright daylight, dim artificial light, abrupt changes near windows). These factors may affect direct methods (DSO) and feature-based methods (ORB-SLAM3) differently. Moreover, DSO includes a photometric calibration module designed to compensate for the camera's response function and vignetting, but its actual effectiveness under real-world indoor lighting variations has not been thoroughly evaluated. Existing comparative studies often rely solely on public datasets with ground truth and neglect the common scenario where ground truth is unavailable. A recent study by Crocetti et al. [5] evaluated DSO and ORB-SLAM3 in low-light indoor environments using robot-mounted auxiliary lighting, but did not investigate natural daylight-to-artificial light transitions or the role of photometric calibration. Thus, there remains a need for a comprehensive evaluation that covers multiple indoor scene types, realistic illumination changes, and both quantitative (ground-truth) and qualitative (no-ground-truth) assessments.

Methodology

This study evaluates two monocular visual SLAM systems: Direct Sparse Odometry (DSO) [3] and ORB-SLAM3 [4]. Both systems are used in their official implementations without functional modifications, except for minor additions to ORB-SLAM3 for logging purposes. All experiments are conducted offline with graphical interfaces disabled to ensure consistency and avoid rendering overhead.

DSO is taken from the official repository released in 2016 [3]. The system provides three operation modes controlled by the mode parameter: mode=0 applies photometric calibration using externally supplied files for camera response function and vignetting; mode=1 performs a simple online photometric compensation without requiring calibration data; and mode=2 disables all photometric processing, using raw pixel values directly. In this work, DSO is executed with the default preset (preset=0), which defines core optimization parameters including a desired point density of 2000, an immature point density of 1500, a sliding window of 5 – 7 keyframes, and up to 6 Levenberg–Marquardt iterations at original image resolution. Geometric calibration is provided via a standard calibration file. For self-recorded sequences, DSO is evaluated in two configurations: with full photometric calibration (mode=0) using calibration files generated by the `online_photometric_calibration` tool, and with online compensation (mode=1). All DSO runs are performed with visualization disabled.

ORB-SLAM3 is used from the official repository released in 2021 [4], running in monocular mode with input adapted from the `mono_tum.cc` example. Default ORB parameters are retained throughout the experiments, including 1000 features per frame, a scale factor of 1.2, and 8 pyramid levels, all specified via YAML configuration files customized for each sequence. To enable a fair comparison with DSO in terms of tracking quality, two minimal modifications were made that do not alter the core SLAM algorithm: an interface was added to expose the number of inlier features in the current frame, and the main loop was extended to write this information to a log file when the corresponding environment variable is set. Additionally, an option was added to disable the Pangolin visualization for headless execution. All ORB-SLAM3 runs are also performed without visualization.

Two categories of image sequences are used in this study. The first category consists of two public benchmark datasets that provide ground truth trajectories, serving as a baseline for performance evaluation. The second category includes four self-recorded sequences captured in real indoor environments without ground truth, designed to investigate the impact of lighting variations and photometric calibration.

The TUM RGB-D `viconroom1` easy sequence [6] is captured in a laboratory environment using an RGB-D camera at 640×480 resolution and 30 frames per second, with ground truth provided by a motion capture system. In this study, only the RGB images are used to simulate monocular input, while depth information is discarded. The EuRoC MAV sequence `V1_01_easy` [7] is collected from a micro aerial vehicle flying through an industrial environment, providing stereo images at 752×480 resolution and 20 frames per second, with ground truth from a laser tracker. Only the left images are used for monocular evaluation.

For the self-recorded sequences, two typical indoor scenes were selected: a university classroom measuring approximately 10 by 8 meters, containing desks, chairs, a blackboard, and windows; and a building corridor approximately 30 meters long and 2.5 meters wide, with white walls and doors. For each scene, two sequences were captured along similar paths under different lighting conditions: daytime with only natural light coming through windows, and nighttime with only artificial ceiling lights, resulting in four sequences in total.

All sequences were recorded using a smartphone rear camera at 480×640 resolution and 30 frames per second, saved in MP4 format. The videos were converted to PNG image sequences using FFmpeg with a quality setting that ensures high output fidelity. Timestamps were generated using OpenCV in Python by obtaining the frame rate and total frame count from each video, then computing timestamps as frame index divided by frame rate. The resulting timestamp files follow the format required by the SLAM systems, with each line containing the frame index, timestamp, and a placeholder for exposure time that is not used in this study.

Geometric camera parameters were obtained using a standard checkerboard pattern and an online OpenCV calibration routine. For DSO experiments requiring photometric calibration, calibration files containing the camera response function and vignetting were generated using the

online_photometric_calibration tool, which processes multiple images captured under varying illumination in each scene. No ground truth trajectories are available for these self-recorded sequences.

To systematically evaluate the performance of the two algorithms in indoor environments, four experiments were designed with different datasets and system configurations. All experiments were conducted offline on the same hardware platform to ensure consistency of results.

First, baseline evaluation is conducted on public benchmarks with metric ground truth. ORB-SLAM3 and DSO are run using the identical code paths and hardware to be adopted for in-house data, and monocular trajectories are aligned with the ground truth references via a similarity transform to standardize scale and pose for comparability. This step validates the expected performance of the local algorithm build and establishes a clear quantitative baseline for the subsequent analysis of self-recorded corridor and classroom sequences without ground truth.

Next, the core study is carried out on four self-recorded sequences from two typical indoor scenes – an open 509 classroom loop and a narrow corridor, each captured under daytime and nighttime lighting conditions. ORB-SLAM3 and DSO (in its calibrated and self-compensated photometric modes) are executed on each sequence, and in the absence of ground truth, absolute trajectory error is not reported. Instead, tabulated run statistics are summarized, estimated paths are overlaid for qualitative.

Comparison, and instances of tracking breaks or multi-map fragments are recorded where they appear. After presenting the trajectory and tabular results, computational efficiency is analyzed for all test conditions. For every sequence and algorithm configuration, the mean tracking time per frame is aggregated from repeated runs with an unchanged logging protocol. These efficiency metrics are presented alongside the qualitative findings, so that the denser direct mapping and more computationally expensive optimization of direct methods are not misconstrued as cost-free gains in scenarios where robustness is similar or inferior.

Finally, a purely visual comparison of scene reconstruction quality is performed. Representative viewpoints where both systems maintain stable tracking are selected, and the reconstruction outputs of the two methods are compared directly – ORB-SLAM3’s sparse scatter of landmarks against DSO’s semi-dense intensity-bearing point cloud. This step addresses practical deployability requirements, as it allows human operators to judge the usability of the reconstructed geometry and texture in cases where metric scoring cannot be computed.

For the public datasets (TUM and EuRoC), Absolute Trajectory Error (ATE) and Relative Pose Error (RPE) are used as quantitative metrics. Since monocular SLAM suffers from scale ambiguity, the estimated trajectory is first aligned with the ground truth using a 7 degree of freedom similarity transformation (Sim(3)) to compensate for scale drift.

APE (Absolute Pose Error) measures the position deviation between the estimated trajectory and the ground-truth trajectory at each time step. It first aligns the estimated poses using a similarity transformation with scale s , rotation matrix R , and translation vector t , then computes the Euclidean distance between the transformed estimated points and the ground-truth points. Its calculation formula is given by

$$APE_i^{trans} = \|sRp_i + t - p_i^*\|. \quad (1)$$

Max APE (Maximum Absolute Pose Error) measures the position deviation of the frame with the largest error in the entire trajectory. It takes the maximum value among the translational absolute pose errors of all frames, reflecting the worst positioning accuracy during the algorithm execution. Its calculation formula is given by

$$\max APE = \max_i APE_i^{trans}. \quad (2)$$

ATE (Absolute Trajectory Error) performs root-mean-square statistics on the absolute pose errors over the entire trajectory. It first computes the translational absolute pose error for each frame, then averages the squared errors across all frames and takes the square root, reflecting the overall positioning accuracy of the algorithm. Its calculation formula is given by

$$ATE = \sqrt{\frac{1}{N} \sum_{i=1}^N (APE_i^{trans})^2}. \quad (3)$$

RPE (Relative Pose Error) evaluates the estimation accuracy of motion between consecutive local frames. It first computes the relative transformations between adjacent frames for both ground-truth and estimated poses, then derives the translational or rotational deviation from the error matrix, and takes its norm as the per-frame relative error. Its calculation formula is given by

$$E_i^{rel} = (\Delta T_i^{gt})^{-1} \Delta T_i^{est}, \quad e_i^{RPE} = \|\tau(E_i^{rel})\|, \quad RPE_{RMSE} = \sqrt{\frac{1}{M} \sum_i (e_i^{RPE})^2}. \quad (4)$$

For the self-recorded sequences (classroom and corridor, day and night), which lack ground truth, a combination of qualitative and semi-quantitative measures is employed. The evaluation focuses on the following aspects: tracking continuity, i.e., whether the system maintains tracking throughout the entire sequence; if failures occur, the number of losses and successful relocalizations are recorded; trajectory smoothness is inspected visually for abrupt jumps or jitter; for sequences with loop closure, the Euclidean distance between the start and end points after aligning the start point is used as a rough estimate of drift; per-frame statistics, including the number of optimized points for DSO and the number of inlier features for ORB-SLAM3, are logged to reflect tracking quality under varying illumination; and average processing time per frame (in milliseconds) is measured to compare computational efficiency.

These metrics allow a comprehensive comparison of the two algorithms from the perspectives of localization accuracy, robustness, and real-time performance.

Experimental Results and Analysis

This chapter conducts benchmark tests based on TUM and EuRoC public datasets (5 repeated runs, official algorithm implementations with only log/visualization adjustments) to verify the stability of the experimental environment and establish a reference for subsequent analysis of self-recorded datasets. The impacts of illumination, texture, and photometric calibration on algorithm performance will be further explored using self-recorded indoor scene datasets.

Table 1 summarizes the 5-time average measured performance of the two algorithms in terms of accuracy, efficiency, mapping, and other dimensions, while marking the reference values published in official papers/benchmark libraries to achieve direct comparison between measured results and official data.

Table 1. **Five-time Average Measured Performance of DSO and ORB-SLAM3 on Public Datasets**

Algorithm	Test Dataset	ATE(m)	RPE(m)	Max APE (m)	Single Frame Time (ms)	Resolution
DSO mode0	TUM Sequence 14	0.2322	0.6016	0.4862	20.05	1280×1024 /640×480
ORB-SLAM3	EuRoC V1_01_easy	0.0875	0.0271	0.1560	11.51	752×480 /600×350

These two trajectory (see Figure 1 and Figure 2) plots were produced on public datasets and show monocular estimates compared to ground truth after Sim(3) alignment (scale and pose). They allow a direct visual check of our local implementation and runtime behavior: ORB-SLAM3 stays closer to the ground truth and yields a smoother path, whereas DSO follows a similar overall shape but exhibits greater drift and local fluctuations.

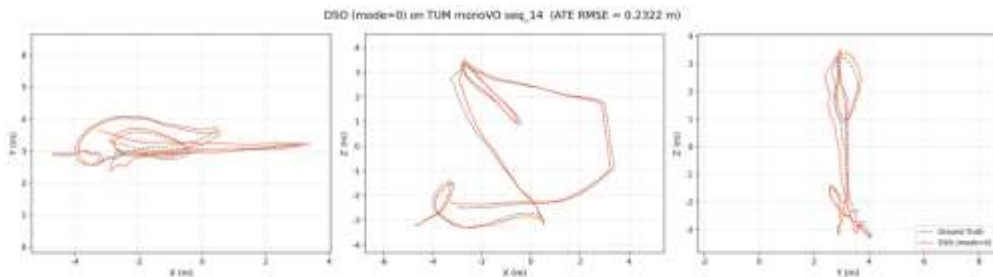


Figure 1. 3D Trajectory Comparison of ORB-SLAM3 and DSO (mode1) on TUM freiburg1_xyz

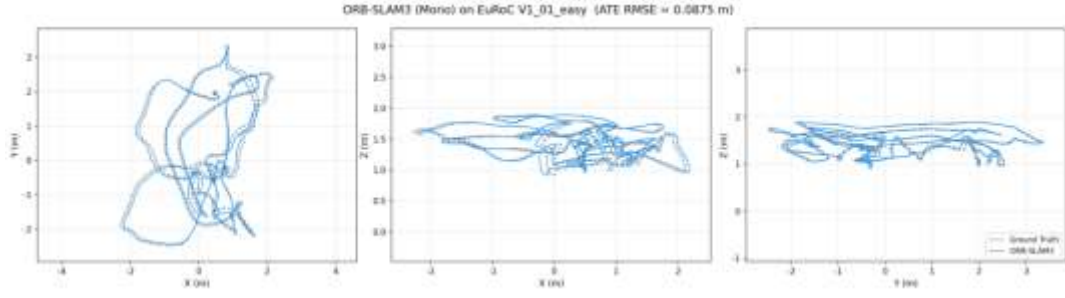


Figure 2. 3D Trajectory Comparison of ORB-SLAM3 and DSO (mode1) on EuRoC V1_01_easy Dataset

This section presents the experimental results on four self-recorded sequences (see Figure 3), designed to evaluate the performance of ORB-SLAM3 and DSO (mode0/mode1) under realistic illumination variations and different scene textures. All sequences were captured in typical indoor environments, with consistent camera motion paths and hardware/software configurations to ensure comparability.



Figure 3. Representative Frames of Four Self-Recorded Test Sequences: *a* – 509 day; *b* – 509 night; *c* – corridor day; *d* – corridor night

Data were recorded in two representative indoor settings: Classroom 509 and a narrow corridor. 509 by day is relatively bright but spatially non-uniform because of windows and exterior light. The room contains many objects, yet desks and chairs have glossy, weakly textured surfaces. The route deliberately includes a segment looking toward a window, creating strong highlights and a high dynamic range that stress both feature extraction and photometric modeling. 509 at night follows roughly the same path under weaker artificial light or mixed outdoor illumination, testing behavior as illumination and contrast fall.

The corridor by day is geometrically confined with fewer static objects, but artificial lighting is more even. The camera moves mainly back and forth along a fixed direction, producing regular field-of-view changes and a controlled “straight–turn” motion pattern for assessing tracking continuity. The corridor at night lowers ambient brightness in the same layout to see whether features and image gradients remain usable. Floors in both environments show irregular wear, stains, and joints that provide low-level texture for visual odometry, but they do not remove differences driven by scene structure and lighting.

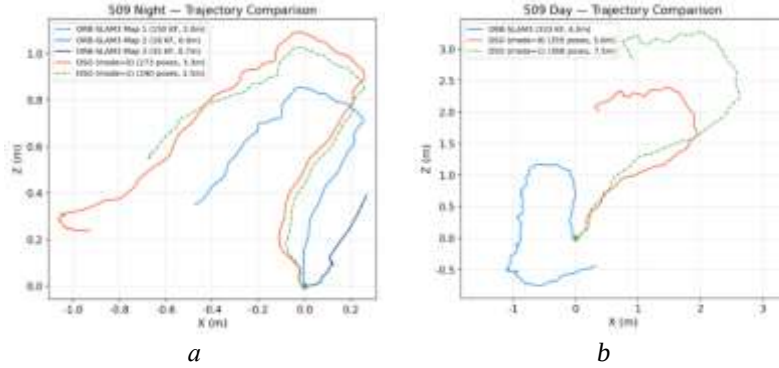
Together, the four sequences contrast illumination uniformity, semantic clutter, specular/window glare, and motion pattern: 509 emphasizes rich clutter and non-uniform lighting; the corridor emphasizes more uniform illumination, repetitive geometry, and constrained heading. The day – night pairs systematically isolate changes in illumination within the same space for comparing ORB-SLAM3 and DSO.

Given nearly identical input lengths on the Classroom 509 sequence for day and night across the three setups (ORB-SLAM3, DSO mode = 0 and mode = 1), DSO produces denser pose outputs and roughly an order of magnitude more observations per frame than ORB-SLAM3, consistent with dense photometric sampling versus sparse feature matching (see Table 2). All metrics drop at night relative to day, with a steeper decline for ORB-SLAM3. The two DSO modes remain close in value, differing only slightly between day and night.

Table 2. **Five-time Average Measured Performance on Classroom 509 Sequence**

Metric	ORB-SLAM3 (Day/Night)	DSO mode=0 (Day/Night)	DSO mode=1 (Day/Night)
Input Frames	2302 / 1551	2301 / 1550	2301 / 1550
Output Poses	322 / 201	352 / 268	362 / 272
Average Points per Frame	252.1 / 218.7	2404.9 / 2065.1	2436.1 / 2035.0

On our self-recorded 509 closed-loop path (see Figure 4), all three configurations produce a recognizable looping motion under daytime illumination. ORB-SLAM3 yields a trajectory that better resembles a closed loop, whereas DSO (mode = 0 and mode = 1) is more prone to map loss during segments where the camera is deliberately pointed at bright windows. Because relocalization is unavailable, lost maps cannot be recovered into a single consistent reference, so the degree of loop closure is generally weaker than that of ORB-SLAM3. The estimates also do not coincide in global shape, reflecting the intrinsic scale and coordinate-frame ambiguity of monocular vision. At night, ORB-SLAM3 undergoes repeated map resets, fragmenting the estimate into multiple segments in different local frames, which cannot be read as one continuous closed path on a single plot; by contrast, both DSO modes maintain one continuous trajectory from the start. Overall, low light accentuates the instability of feature-based SLAM under multi-map switching, while the direct formulation tends to be more robust in terms of trajectory continuity. Importantly, without ground truth and without loop closure, these curves should be treated only as qualitative comparisons of shape and continuity, and must not be interpreted as metrically accurate scale or strict loop-closure performance for monocular estimates.

Figure 4. Trajectory Evaluation of ORB-SLAM3 and DSO in 509 Room: *a* – 509 night; *b* – 509 day

The corridor sequence involves a narrow layout and a controlled walking pattern. The table shows that ORB-SLAM3’s number of output poses decreases at night, yet it does not exhibit the same kind of multi-map fragmentation seen on 509 at night after repeated resets. Average points per frame remain about an order of magnitude higher for DSO than for ORB. DSO’s point density changes only mildly between day and night in the corridor, which suggests that under relatively uniform artificial lighting the scale of direct-method sampling stays stable; ORB’s points per frame are also similar between day and night (see Table 3).

Table 3. **Five-time Average Measured Performance on Corridor Sequence**

Metric	ORB-SLAM3 (Day/Night)	DSO mode=0 (Day/Night)	DSO mode=1 (Day/Night)
Input Frames	1805 / 1861	1804 / 1860	1804 / 1860
Output Poses	248 / 206	279 / 208	289 / 212
Average Points per Frame	212.4 / 214.3	2639.8 / 2631.0	2707.2 / 2667.9

Day and night, the corridor trajectories exhibit very small lateral motion (see Figure 5), with the bulk of the motion forming an approximately straight path. By day, the three estimates extend farther

along the vertical axis and stay closer to each other. At night, the overall extent along that axis is shorter and the three traces separate somewhat more, yet the pattern remains a near-straight line dominated by forward motion, without the clear multi-map, piecewise structure seen for ORB-SLAM3 at night on the 509 sequence.

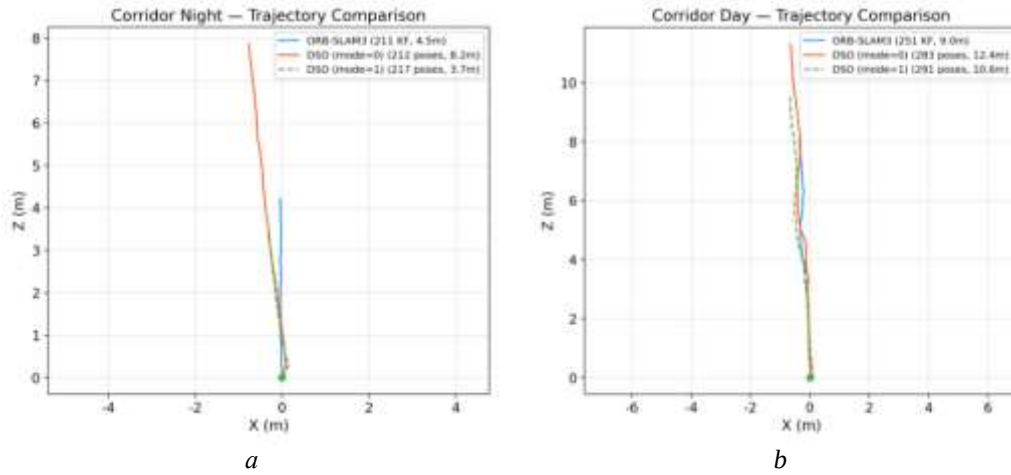


Figure 5. Trajectory Performance Comparison of ORB-SLAM3 and DSO in Corridor Scenarios Under Day and Night Illumination *a* – corridor night; *b* – corridor day

To further characterize the computational cost of our self-collected sequences, we report the average per-frame tracking time of ORB-SLAM3 and DSO (mode = 0 / mode = 1) on the 509 and corridor scenarios under day and night illumination (see Figure 6). The statistics show that ORB-SLAM3 remains near 8.3 – 9.2 ms per frame in all four conditions and is substantially faster than DSO; both DSO configurations span 19.2 – 28.9 ms per frame, i.e., roughly 2.5 – 3 $\times$ the ORB-SLAM3 timings. This gap is consistent with the differing computational structures: DSO forms photometric residuals over many high-gradient samples inside a sliding window and performs iterative optimization, yielding a larger per-frame problem; ORB-SLAM3 relies mainly on feature extraction, matching, and comparatively light backend updates, and therefore exhibits lower time cost on our hardware and image-resolution settings.

Within DSO, mode = 1 is never faster than mode = 0 across all scenes; the largest gap appears on the corridor day sequence (≈ 25.0 ms/frame vs. 28.9 ms/frame), suggesting that disabling fixed photometric calibration and relying on online affine brightness compensation may increase the number of optimization variables or iterations and thus raise the average latency. Day–night changes barely affect ORB-SLAM3, whose curve is almost flat; DSO decreases to some extent at night on both 509 and the corridor, with a more pronounced drop on corridor night. This may reflect fewer reliable gradients under low light, changes in keyframe triggering, or a smaller active set of residuals in the window, shrinking each optimization step and lowering the per-frame average. Lower latency must not be equated with higher accuracy; interpretation should be combined with trajectory continuity, mapping density, and failure rates reported elsewhere.

Overall, the figure quantifies – in terms of wall-clock cost – the price of denser geometric and photometric modeling: ORB-SLAM3 is more lightweight per frame, whereas DSO trades higher per-frame cost for richer direct-map information, with mode = 0 slightly favoring average efficiency over mode = 1.

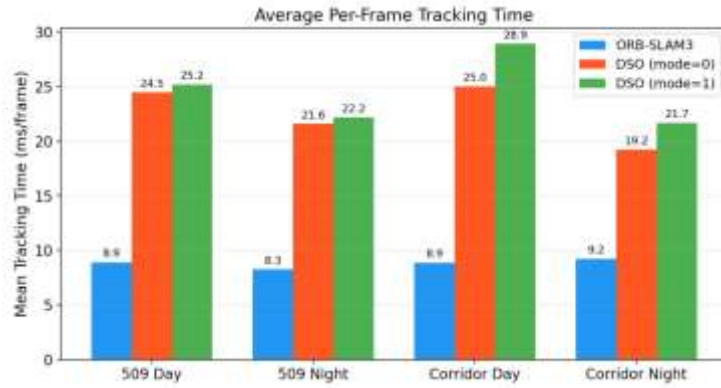


Figure 6. Average Per-Frame Tracking Time of ORB-SLAM3 and DSO Under Different Scenarios

DSO minimizes photometric residuals over many high-gradient samples in a sliding window and publishes a semi-dense map with intensity, so points tend to form spatially coherent patches; floor patterns, wall boundaries, and furniture silhouettes are easier to recognize. On the right, the bench region accumulates enough projected points that a human-shaped gray clutter emerges – this reflects coupled photometric–geometric constraints rather than semantic understanding or true dense reconstruction. ORB-SLAM3 builds a sparse landmark map tied to detected features; smooth or weakly textured surfaces (bench, clothing, uniform walls) yield few stable landmarks, leaving only scattered red dots where corners exist. Thus, for this viewpoint and corridor, DSO’s map is far more interpretable by eye, while ORB-SLAM3’s map remains a sparse geometric scaffold (see Figure 7).

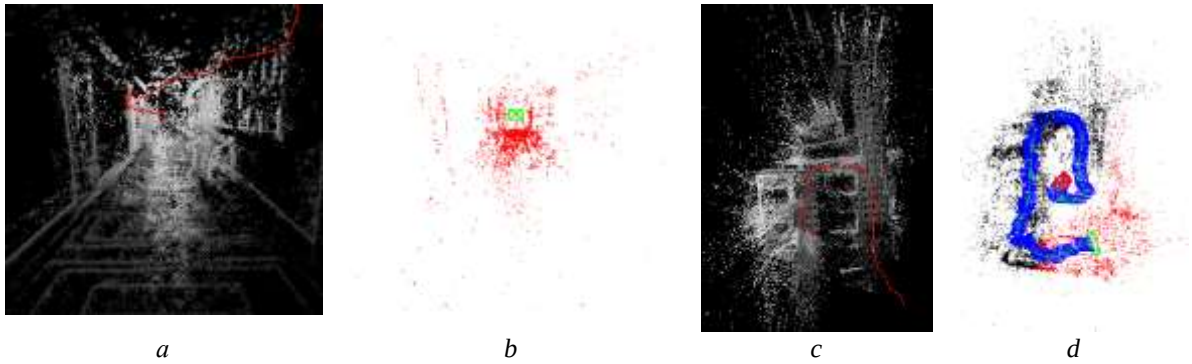


Figure 7. Reconstructed Scene Texture: A Comparison of Different SLAM Algorithms: *a* – DSO corridor; *b* – ORB corridor; *c* – DSO classroom; *d* – ORB classroom

Conclusion

The experiments first complete a baseline validation of the local implementation on public datasets, then turn to self-recorded sequences. This order gives the subsequent claims a clear traceability: the public benchmarks check that the pipeline, parameters, and runtime environment are sound; the custom sequences expose how the algorithms behave in realistic indoor settings. Following this protocol, the paper compares methods along four dimensions – trajectory shape, summary statistics, average per-frame time, and mapping visualization – to form a relatively complete evaluation framework.

On the self-recorded data, ORB-SLAM3 and DSO show a fairly clear division of roles. ORB-SLAM3 is overall stronger in pose estimation and trajectory stability: when illumination and texture are favorable, the estimated path aligns better with the actual motion and global consistency is easier to maintain. DSO tends to produce visually richer maps with more structural detail; its semi-dense output makes scene contours and local texture easier to perceive. Neither is free of failure modes: ORB-SLAM3 is more prone to tracking loss and map resets under night conditions and strong specular highlights; without loop closure, DSO can drift more noticeably and may also fail when facing very bright regions. The paper therefore does not frame the comparison as a simple winner – loser ranking, but as complementary strengths in localization robustness versus map interpretability.

In terms of computational cost, ORB-SLAM3's average per-frame tracking time is lower than DSO's throughout, consistent with the different front-end workload and optimization scales of feature-based and direct formulations. That suggests ORB-SLAM3 is more likely to meet tight latency budgets, whereas DSO trades higher per-frame cost for denser scene expression. Taken together with the trajectory and mapping results, system design should trade off localization stability, runtime, and map detail in a task-driven way rather than relying on a single score.

The conclusions remain subject to clear experimental limits. First, no standard motion-capture or survey-grade ground truth was collected for the custom sequences, so ATE, RPE, and similar absolute metrics are not reported there; the analysis is mainly comparative and phenomenological. Second, only five runs were performed per condition, which is a modest sample for highly stochastic monocular systems and limits statistical generality. Third, the 509 recording includes deliberate camera motion toward bright windows, and the session does not end at the same pose as the opening frame; both choices increase the difficulty of loop closure and weaken conditions for detectors such as those in ORB-SLAM3, so some effects reflect the acquisition protocol as well as the algorithms. For these reasons the results are best read as behavioral analysis rather than rigorous metric benchmarking.

Building on these observations, future work will proceed in three directions: broaden coverage with more scene types and lighting layouts and scale repetition to on the order of twenty runs, reporting richer summary statistics; where possible add usable ground truth or auxiliary metrology so custom data can be folded into standard error metrics; and explore fusion architectures that use ORB-SLAM3 for global stability while leveraging DSO's semi-dense direct mapping to improve map legibility. Overall, this study not only contrasts the two approaches in real indoor conditions but also clarifies the problem boundaries and design space for subsequent hybrid visual SLAM / VO systems.

References

1. Cadena C., Carlone L., Carrillo H. et al. // IEEE Transactions on Robotics. 2016. Vol. 32. P. 1309–1332.
2. Mur-Artal R., Tardós J. D. // IEEE Transactions on Robotics. 2017. Vol. 33. P. 1067–1080.
3. Engel J., Koltun V., Cremers D. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018. Vol. 40. P. 184–197.
4. Mur-Artal R., Campos C. // IEEE Transactions on Robotics. 2021. Vol. 37. P. 1874–1890.
5. Crocetti F., Brilli R., Dionigi A. et al. // Journal of Field Robotics. 2025. Vol. 42. P. 3748–3771.
6. Sturm J., Engelhard N., Endres F. et al. // Proc. of the 2012 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. 2012. P. 573–580.
7. Burri M., Nikolic J., Gohl P. et al. // Proc. of the 2016 IEEE Int. Conf. on Intelligent Robots and Systems. 2016. P. 1547–1554.

UAV-TO-SATELLITE IMAGE GEO-LOCALIZATION VIA DUAL-BRANCH DEEP LEARNING: A COMPARATIVE STUDY OF RESNET-18 AND MLP-MIXER

QINGHAN YU

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus**Received April 5, 2026*

Abstract. Unmanned Aerial Vehicle (UAV)-satellite cross-view geolocalization is a core technology for autonomous navigation in GNSS-denied environments, which achieves precise pose estimation by matching low-altitude oblique UAV images with high-altitude orthophoto satellite images. However, the huge visual differences in perspective, scale and distortion between the two types of images significantly increase the difficulty of feature matching. Current mainstream methods are mostly based on CNN or Transformer architectures, while systematic comparative studies of pure MLP architectures in this task are still insufficient, and their performance differences and efficiency trade-offs remain unclear. To this end, this paper compares and analyzes the accuracy, robustness and computational efficiency of CNN and pure MLP architectures in UAV-satellite cross-view geolocalization on a unified benchmark dataset. The experimental results show that CNN has advantages in local feature extraction and is more robust to perspective distortion, while pure MLP architectures show potential in global feature modeling and inference speed, but their performance needs to be improved in small-sample and extreme perspective change scenarios. The comparative analysis in this paper provides an empirical basis for the architecture selection and optimization of cross-view geolocalization models.

Keywords: cross-view geolocalization; UAV-satellite images; CNN; pure MLP architecture; feature matching.

Introduction

Unmanned Aerial Vehicles (UAVs) have been widely applied in key scenarios such as autonomous navigation, emergency search and rescue, and land resource surveying due to their advantages of flexible deployment, low cost, and high-resolution imaging, among which UAV-satellite cross-view geolocalization serves as one of the core technologies supporting these tasks, achieving precise pose estimation of UAVs in the global geographic coordinate system by matching low-altitude oblique UAV images with high-altitude orthophoto satellite images, thus providing a reliable solution for positioning in GNSS-denied environments [1]. However, there exist huge visual differences between UAV and satellite perspectives, including perspective and scale differences where UAVs capture low-altitude oblique images with flexible and variable viewpoints while satellites take high altitude orthophotos with fixed perspectives, leading to distinct visual appearances of the same geographic entity due to significant disparities in imaging scale and projection mode, rotation and distortion differences caused by unstable UAV flight attitudes that introduce image rotation and distortion to further increase feature matching difficulty, as well as illumination and seasonal differences that alter lighting conditions and vegetation coverage to affect feature stability and repeatability, all of which pose severe challenges to cross-view feature matching [1]. At present, research on cross-view geolocalization mainly focuses on CNN and Transformer architectures, with the former performing well in feature extraction by virtue of local inductive bias and the latter excelling in global context modeling. Yet pure MLP architectures, as a simple model without inductive bias that has shown performance comparable to CNN in general computer vision tasks [5], have not been verified for their applicability in UAV-satellite cross-view geolocalization, and there remains a clear gap in systematic comparative studies with CNN, leaving their performance differences and efficiency trade-offs in this task unclear. Against this background, this

paper conducts a systematic comparative experiment between CNN and pure MLP architectures in UAV-satellite cross-view geolocalization on a unified benchmark dataset [4], analyzing their differences in positioning accuracy, robustness to visual variations, and computational efficiency, revealing the performance rules of different architectures under key challenges such as perspective distortion and scale change to provide empirical guidance for model selection, and discussing the optimization direction of pure MLP architectures in this scenario to offer a reference for subsequent research, with the rest of the paper organized as follows: Section 2 reviews related work on cross-view geolocalization and pure MLP architectures, Section 3 introduces the experimental dataset, comparative models, and evaluation metrics, Section 4 analyzes and discusses the experimental results, and Section 5 summarizes the full paper and prospects future work.

Related Work

Cross-view geolocalization focuses on matching low-altitude UAV oblique images with high-altitude satellite orthophotos to estimate precise UAV poses in GNSS-denied environments, serving as a core technology for autonomous navigation and emergency search and rescue [1]. Classic benchmark datasets including CVUSA [6], VIGOR [7], and University-1652 [3] have been established to evaluate task performance, among which University-1652 is widely adopted as the standard testbed for current UAV-satellite cross-view localization due to its diverse campus and urban scenes and well-aligned matching image pairs [3]. Early studies relied on hand-crafted features such as SIFT [8] and HOG [9] to capture invariant visual patterns for cross-view matching, but these methods struggled with severe perspective, scale, and distortion differences between UAV and satellite images, resulting in limited accuracy and robustness that failed to meet practical application demands. With the rise of deep learning, data-driven approaches gradually replaced hand-crafted feature methods, opening new avenues for addressing the challenges of cross-view geolocalization.

CNN-based architectures have dominated cross-view geolocalization research due to their inherent local inductive bias, which enables effective extraction of key local visual features such as building edges, road intersections, and texture patterns that are critical for matching. Representative works include NetVLAD [10] and CVM-Net [1]: NetVLAD introduces a vector of locally aggregated descriptors to aggregate convolutional features into compact global representations, laying the foundation for efficient large-scale cross-view image retrieval [10], while CVM-Net employs a siamese network structure with shared CNN weights to learn shared feature spaces between UAV and satellite images, and uses metric learning objectives to minimize the feature distance of matching image pairs and maximize that of non-matching pairs, significantly improving cross-view matching accuracy [1]. Despite their advantages in local feature robustness and parameter efficiency, CNN-based methods suffer from limited global context modeling ability, as their fixed-size receptive fields make it difficult to capture long-range dependencies between geographic entities, leading to potential feature misalignment under extreme perspective and scale variations.

In recent years, Transformer architectures have emerged as promising alternatives to CNNs in computer vision [11], offering new possibilities for cross-view geolocalization. TransGeo applies the self-attention mechanism of Transformers to cross-view localization, modeling global contextual relationships between image regions to alleviate the limitations of CNNs' local receptive fields, thus achieving better performance in scenarios with large-scale perspective differences [2]. As a representative pure MLP architecture, MLP-Mixer discards convolutional and attention operations entirely, relying solely on fully connected layers to interact across spatial and channel dimensions of feature maps, and has achieved performance comparable to state-of-the-art CNNs on general visual benchmarks like ImageNet, demonstrating its potential in simple structure and global feature modeling [5]. However, existing studies in cross-view geolocalization mainly focus on Transformer-based models, and the applicability of pure MLP architectures in UAV-satellite extreme perspective difference scenarios remains unvalidated, with a clear lack of systematic comparative analysis between pure MLPs and classic CNNs in this specific task.

Metric learning losses play a crucial role in optimizing cross-view geolocalization models, as the task essentially boils down to image retrieval where the goal is to rank matching image pairs higher than non-matching ones. Early works predominantly used Triplet Loss [12], which constructs triplets consisting of an anchor image, a positive matching image, and a negative non-matching image to pull

the anchor and positive features closer while pushing the anchor and negative features apart, effectively optimizing the feature embedding space for cross-view matching. Subsequent studies have shifted to contrastive learning losses such as InfoNCE and NT-Xent [13], which sample negative examples within a training batch to construct more efficient contrastive constraints, leading to faster convergence and better generalization performance under large-batch training settings. This loss has become a standard choice for training current cross-view geolocalization models [14]. Despite the maturity of metric learning loss designs, their compatibility and performance trade-offs with different backbone architectures in cross-view geolocalization have not been thoroughly explored.

Methodology

This work focuses on cross-view geo-localization, which aims to find the satellite image tile that matches a given UAV query image in geographic location. Each matching pair of UAV and satellite images shares the same GPS coordinates, but differs greatly in viewing angle, imaging scale, and data modality. This task is treated as an image retrieval problem, where the model learns an embedding function to make matching image pairs more similar and non-matching pairs more distinct in the feature embedding space [15]. For each UAV query image with a known correct satellite match, the model is trained to ensure that the similarity between the query and its ground-truth match is higher than the similarity between the query and any other non-matching satellite tile. The similarity measurement used in this work is cosine similarity, and all feature embeddings are normalized using L2 normalization to ensure stable and consistent similarity calculation [14].

The proposed framework uses a dual-branch structure with two independent encoders for UAV images and satellite images respectively. The design of independent branches is based on the large domain gap between UAV and satellite imagery. Using shared weights for both branches would force the two different types of images into a single feature space, which is proven to be ineffective through experiments. For each pair of matching UAV and satellite images, the two branches extract fixed-dimensional feature embeddings independently. These embeddings are then normalized and used for contrastive learning. The dual-branch structure allows each branch to learn specialized features suitable for its own image type: UAV images contain rich texture details from oblique perspectives, while satellite images show clear global spatial layouts from top-down views. Through contrastive learning, the two sets of features are aligned to form a unified embedding space where cross-view matching can be effectively realized [1].

ResNet-18 is a lightweight convolutional neural network with residual connections, containing four residual stages with gradually increasing channel dimensions [16]. For an input image, the network extracts features layer by layer and outputs a fixed-length feature vector through global average pooling. Both the UAV branch and the satellite branch are initialized with weights pre-trained on the ImageNet dataset [17], which is critical for model performance as verified in the ablation study. Residual connections ensure smooth gradient propagation during training [16], and hierarchical feature extraction captures multi-scale spatial information. Each ResNet-18 branch has about 11.2 million trainable parameters. The choice of ResNet-18 instead of deeper networks is based on computational efficiency and better generalization to domain-shifted data such as satellite images.

MLP-Mixer is a pure MLP architecture that replaces convolution and self-attention with two alternating MLP operations applied to image patches [5]. The input image is divided into multiple non-overlapping patches, and each patch is projected into a feature token to form a token matrix. Each MLP-Mixer layer executes two operations in sequence: token-mixing MLP for spatial interaction between different patches, and channel-mixing MLP for feature interaction within each patch [5]. The activation function used is GELU, and layer normalization is applied before each MLP operation. The final feature representation is obtained by global average pooling on the token dimension. Unlike CNNs that capture local spatial features layer by layer, MLP-Mixer performs global spatial reasoning from the first layer through token-mixing, which is beneficial for cross-view matching that relies on global landmark layout [5]. The MLP-Mixer branch has about 59.9 million trainable parameters. Its pure MLP structure removes the inductive bias of convolution, forcing the model to learn spatial relationships explicitly, which helps adapt to large viewpoint changes in cross-view tasks.

Following the design of SimCLR [13], a two-layer MLP projection head is added after each backbone network to map backbone features into a compact 256-dimensional embedding space. The

projection head separates the backbone feature learning from the metric learning objective, avoiding over-specialization of backbone features to the contrastive loss [13]. All final embeddings are processed with L2 normalization, so that all embeddings are distributed on a unit hypersphere, making cosine similarity equivalent to dot product similarity [14]. This normalization step is essential for contrastive learning, as it eliminates the influence of embedding magnitude on similarity calculation. The 256-dimensional embedding space is a standard setting that balances feature expression ability and computational cost.

The InfoNCE loss, also known as NT-Xent loss, is used as the main training objective [13]. For a batch of matching UAV-satellite pairs, the loss calculates the contrastive constraint from UAV embeddings to satellite embeddings. A temperature hyperparameter is used to adjust the concentration of the similarity distribution, where a smaller temperature makes matching pairs cluster more tightly [13]. To utilize the symmetry of the cross-view matching task, the loss is applied in both directions: from UAV to satellite and from satellite to UAV. This bidirectional loss optimizes both matching directions simultaneously, improving model convergence and final performance. Compared with Triplet Loss [12], InfoNCE uses all sample pairs in the batch to calculate gradients, providing denser and more effective learning signals [14]. The model learns to maximize the similarity between each query and its correct match, while minimizing the similarity between the query and all other non-matching samples.

In the inference stage, all satellite tiles in the gallery are encoded by the satellite branch to build a feature bank. For each input UAV query image, the model extracts its feature embedding and computes the similarity with all embeddings in the satellite feature bank. The satellite tiles are sorted in descending order of similarity, and the GPS coordinates of the top-ranked tile are taken as the predicted location. The retrieval performance is evaluated using Recall@K metrics, which represent the percentage of queries whose correct satellite match appears in the top-K ranked results [3]. The main evaluation metrics are R@1, R@5, and R@10, among which R@1 is the most stringent and practical index for real geo-localization applications.

To verify the advantage of InfoNCE, a Triplet Loss version with hard negative mining is also implemented [12]. For each UAV query, the hardest negative sample is defined as the non-matching satellite tile with the highest similarity to the query. Triplet Loss only uses this single hardest negative for each anchor, resulting in sparse gradient signals [12]. In contrast, InfoNCE uses all samples in the batch as negatives, providing much denser gradients [14]. The ablation experiment shows that InfoNCE achieves significantly better performance than Triplet Loss, especially for MLP-Mixer, which confirms the rationality of choosing InfoNCE as the training loss.

Experiments and Results

We conduct experiments on the UAV-Visloc dataset comprising 768 image pairs, where each pair consists of a UAV query image and a corresponding satellite gallery tile [4]. The dataset is split into 80 % training and 20 % validation using a fixed random seed to ensure reproducibility. All models are trained for 80 epochs with batch size 32 using the AdamW optimizer [18] with learning rate $1e-4$ and weight decay $1e-4$. The embedding dimension is uniformly set to 256 across all experiments. We evaluate retrieval performance using Recall@K, which measures the percentage of queries for which the correct satellite tile appears in the top-K ranked results. All experiments are conducted on an NVIDIA RTX 5090 GPU with 30GB memory.

As shown in Table 1, we present a comprehensive quantitative performance comparison between the two backbone architectures ResNet-18, a classic convolutional neural network, and MLP-Mixer, a pure MLP-based model on the held-out validation dataset, which serves as a rigorous benchmark to evaluate their generalization and cross-view matching capabilities. Across all three core recall metrics, MLP-Mixer consistently outperforms ResNet-18, demonstrating the clear advantage of its architecture for cross-view geo-localization tasks. Most notably, MLP-Mixer achieves a top-1 recall of 74,09 %, which is 1,69 percentage points higher than ResNet-18’s R@1 score of 72,40 %. This performance lead is consistent across all evaluated recall thresholds: at R@5, MLP-Mixer reaches 88,41 % compared to ResNet-18 87,50 %, marking a 0,91 percentage point improvement, while at R@10, MLP-Mixer’s score of 92,45 % surpasses ResNet-18 90,49 % by a notable 1,96 percentage points. The consistent, across-the-board improvement in all recall metrics strongly indicates that MLP-Mixer’s core token-mixing mechanism, which enables holistic global spatial reasoning from the very first network layer, delivers

far more effective cross-view feature alignment than ResNet-18 hierarchical convolutional design. Unlike CNNs, which rely on local receptive fields to extract features in a layer-wise manner, MLP-Mixer directly models long-range spatial relationships across the entire image, allowing it to capture the overall layout of geographic landmarks rather than focusing solely on local texture details sensitive to viewpoint and scale variations. Notably, the performance gap between the two models is most pronounced at R@1, the strictest and most practically relevant metric for real-world geo-localization applications. This significant improvement in top-1 recall demonstrates that MLP-Mixer learns far more discriminative feature embeddings, which are highly effective at distinguishing matching cross-view pairs from non-matching ones, a critical requirement for reliable UAV autonomous navigation systems.

Table 1. **Overall cross-view retrieval performance of ResNet-18 and MLP-Mixer on validation set**

Model	R@1	R@5	R@10
ResNet-18	72,40 %	87,50 %	90,49 %
MLP-Mixer	74,09 %	88,41 %	92,45 %

To validate the necessity of independent branches for each view, we compare the dual-branch architecture against a variant with shared weights, where a single backbone processes both UAV and satellite inputs. This ablation tests whether the large domain gap between aerial and satellite imagery requires separate feature extraction pathways or if a unified representation space is sufficient. As reported in Table 2, the results reveal a dramatic performance degradation for the shared-weight variant. For ResNet-18, R@1 drops from 72,40 % to 49,35 %, a loss of 23,05 percentage points, indicating that forcing both modalities through a single backbone severely compromises the model ability to learn discriminative cross-view features. Interestingly, MLP-Mixer shared-weight variant shows only marginal degradation, suggesting that token-mixing operations may be more robust to domain shift than CNN local convolutions. However, the independent-branch configuration remains superior for MLP-Mixer as well, confirming that dual-branch architecture is essential for this cross-view task. The asymmetric sensitivity between the two architectures highlights that while MLP-Mixer global receptive field provides some inherent robustness to domain variations, explicit architectural separation for different modalities is still beneficial.

Table 2. **Ablation study: Effect of independent vs shared weight branches on R@1**

Model	Independent Branches	Shared Weights	Δ R@1
ResNet-18	72,40 %	49,35 %	23,0 %
MLP-Mixer	74,09 %	72,79 %	1,30 %

We compare InfoNCE loss with Triplet Loss using hard negative mining (margin=0,3) to verify contrastive learning effectiveness for cross-view geo-localization. InfoNCE computes gradients from all $N \times N$ batch pairs, while Triplet Loss uses only one hard negative per anchor, producing sparser gradients. The comparative results are listed in Table 3. For ResNet-18, both losses yield identical 72,40 % R@1, indicating architecture is the main constraint. For MLP-Mixer, InfoNCE reaches 74,09 % R@1, outperforming Triplet Loss by 5,34 percentage points. The temperature $\tau=0,07$ in InfoNCE supports fine-grained similarity calibration, which well matches MLP-Mixer’s global token-mixing mechanism. Models with global receptive fields better utilize dense contrastive gradients, making InfoNCE the optimal loss for this task, especially when combined with MLP-Mixer.

Table 3. **Ablation study: InfoNCE vs Triplet Loss (hard negative mining, margin=0.3) for R@1**

Model	InfoNCE	Triplet Loss	Δ R@1
ResNet-18	72,40 %	72,40 %	0 %
MLP-Mixer	74,09 %	68,75 %	5,34 %

To quantify the contribution of ImageNet pretraining, we train both models from random initialization with all other hyperparameters kept identical. As summarized in Table 4, the results demonstrate that ImageNet pretraining is critical for cross-view geo-localization: ResNet-18 drops 13,81 percentage points in R@1, while MLP-Mixer suffers a catastrophic 54,04 percentage point decline. This dramatic performance degradation reveals that semantic features from ImageNet are directly transferable to UAV and satellite imagery. The asymmetric impact shows pure MLP architectures, lacking convolutional inductive biases, depend far more on transfer learning; CNNs’ spatial locality and hierarchical feature learning offer better robustness to random initialization. These findings underscore

the necessity of ImageNet pretraining and transfer learning for this task, especially for MLP-based models.

Table 4. Ablation study: Effect of ImageNet pretraining on R@1 performance

Model	Pretrained	Random Init	Δ R@1
ResNet-18	72,40 %	58,59 %	13,81 %
MLP-Mixer	74,09 %	20,05 %	54,04 %

This subsection visualizes the top-5 retrieval results for both ResNet-18 and MLP-Mixer on representative queries. As illustrated in Figure 1, green boxes indicate correct matches, while red boxes indicate incorrect matches. This labeling convention makes it straightforward to visually assess each model’s retrieval accuracy at a glance. The visualization reveals several key differences in retrieval behavior. MLP-Mixer consistently places the correct satellite tile at rank 1 or 2, while ResNet-18 sometimes ranks it lower, particularly in challenging scenarios with significant viewpoint variations. Both models struggle with extreme viewpoint differences, but MLP-Mixer recovers better in the top-5 results, suggesting more robust feature representations. A critical observation is that MLP-Mixer token-mixing operations appear to capture global spatial layouts and landmark configurations more effectively than ResNet-18 local convolutional features. For instance, in cases where buildings or road networks form distinctive patterns, MLP-Mixer correctly matches these patterns across views, while ResNet-18 may be distracted by local texture variations or shadows. These qualitative results corroborate the quantitative findings: MLP-Mixer global receptive field from the first layer enables superior cross-view matching by focusing on semantically meaningful, viewpoint-invariant features rather than low-level texture details.

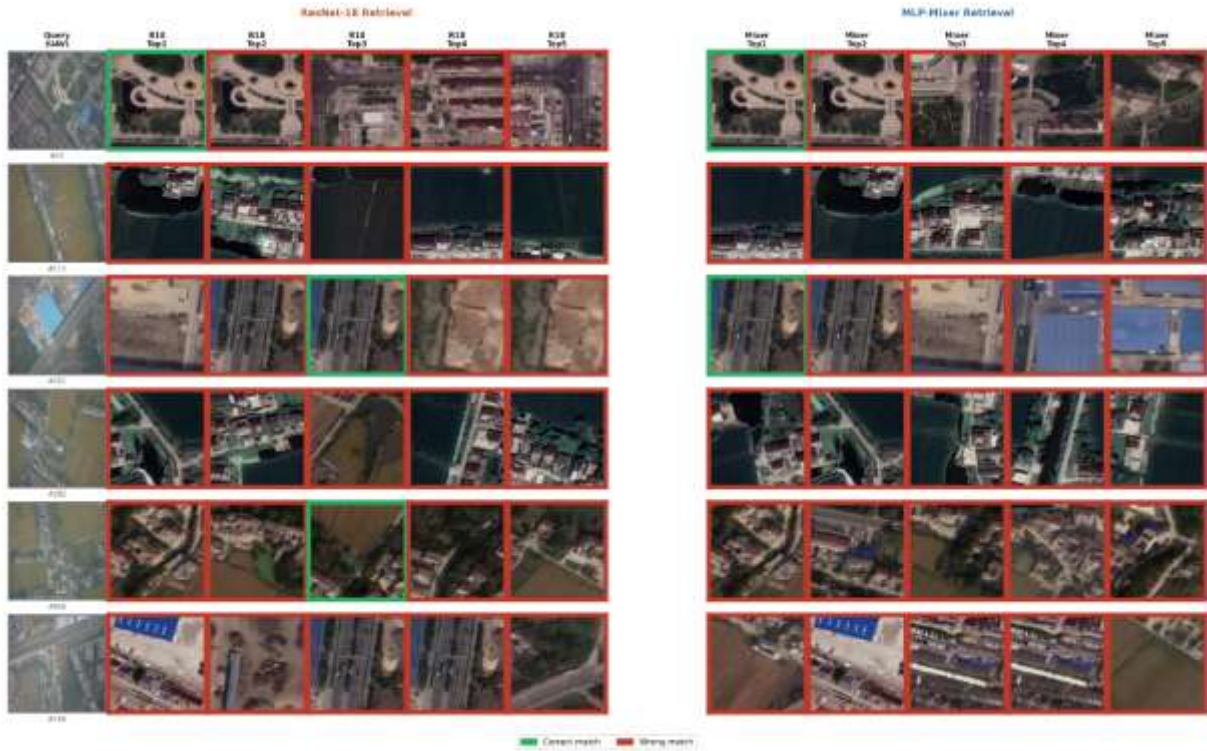


Figure 1. Top-5 UAV-satellite cross-view retrieval results for ResNet-18 and MLP-Mixer. Green boxes denote correct matches, red boxes denote incorrect matches

This subsection visualizes the learned embedding space using t-SNE dimensionality reduction [19]. As displayed in Figure 2, each color represents a unique GPS location. Circles denote UAV query embeddings, and stars denote satellite gallery embeddings. The visualization reveals striking differences in cross-view feature alignment between the two models. For ResNet-18, UAV and satellite embeddings for the same location are scattered across the embedding space, indicating weak cross-view alignment. Many locations show significant spatial separation between circles and stars, suggesting that the model has not learned a unified representation space where matching pairs are close. In contrast, MLP-Mixer demonstrates superior cross-view alignment, with UAV and satellite embeddings clustering tightly

together by location. The clear color-based clustering indicates that the model has learned a unified representation space where matching pairs are proximal in the embedding space. This tight clustering is particularly evident for locations with distinctive landmark configurations, where both UAV and satellite views converge to similar embedding regions. The superior alignment in MLP-Mixer embedding space directly translates to better retrieval performance, as queries and gallery items from the same location are naturally closer in the learned metric space. This visualization provides compelling evidence that MLP-Mixer token-mixing mechanism produces more aligned cross-view embeddings through its global spatial reasoning capabilities.

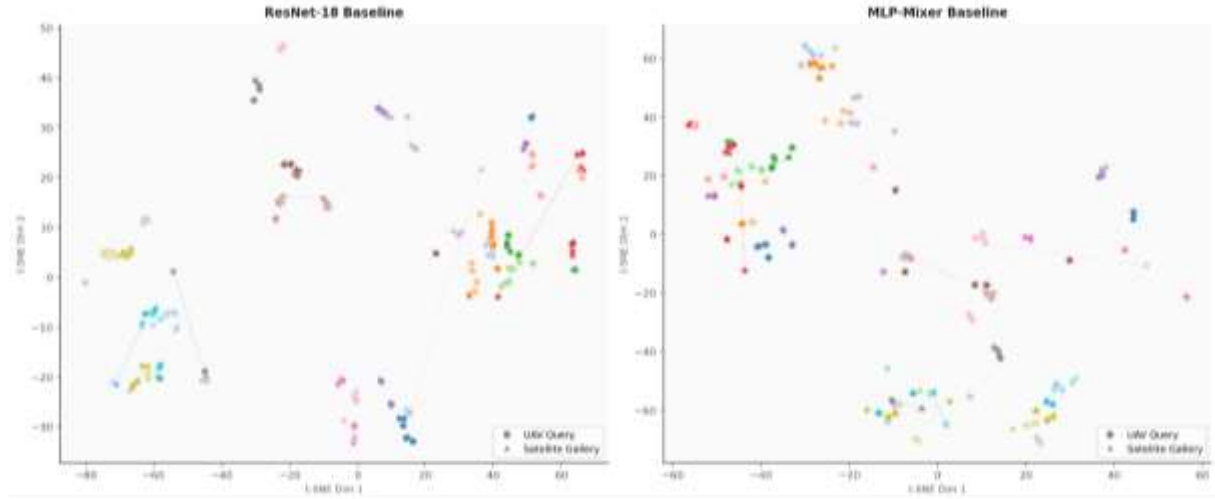


Figure 2. t-SNE visualization of learned cross-view embedding spaces for ResNet-18 and MLP-Mixer. Each color represents a unique GPS location

This subsection presents visualizations of the attention regions for both models using the Grad-CAM method [12], aiming to reveal which image regions contribute most to the generation of the final feature embedding and to deeply analyze the feature extraction mechanism of each network. As demonstrated in Figure 3, the generated attention heatmaps offer clear and intuitive insights into the interpretability of the model decision-making process and the robustness of feature representation, which is crucial for understanding the performance differences between the two architectures in cross-view geo-localization tasks. For ResNet-18, the attention distribution is scattered across multiple irrelevant regions of the image. Instead of focusing on core semantic landmarks that are stable across views, the model often concentrates on low-level textures, local shadows, noise, or non-essential background elements that are highly sensitive to perspective changes. This scattered and unstable attention pattern strongly indicates that ResNet-18 tends to rely on texture-based shallow features, which are vulnerable to variations in viewing angle, imaging scale, and illumination, thus leading to poor feature stability in cross-view matching. In contrast, MLP-Mixer presents highly concentrated attention on high-level semantic structural features, including typical buildings, complete road networks, regular vegetation distribution patterns, and uniquely identifiable landmark configurations. These semantic features possess strong invariance across oblique UAV views and orthogonal satellite views, making them ideal cues for cross-view matching. More importantly, MLP-Mixer attention maps maintain highly consistent focus regions for corresponding UAV and satellite image pairs, which directly proves that the model has successfully learned viewpoint-invariant semantic features during the training process. For instance, when a prominent building or a critical road intersection exists in both the UAV image and the satellite image, MLP-Mixer stably focuses on these key regions in both views, while ResNet-18 attention shifts randomly and fails to align across views. This excellent cross-view attention consistency is the key reason for MLP-Mixer superior retrieval performance: by capturing semantically meaningful and perspective-insensitive features, the model constructs highly robust feature representations that can adapt to the drastic perspective differences inherent in UAV-satellite cross-view geo-localization. Benefiting from the global token-mixing operations inside the architecture, MLP-Mixer is capable of reasoning about global spatial relationships and overall landmark configurations in a holistic manner. Unlike ResNet-18, which is easily disturbed by local texture fluctuations and view-specific imaging

artifacts, MLP-Mixer maintains stable and focused feature perception, further enhancing its reliability and accuracy in cross-view matching tasks.

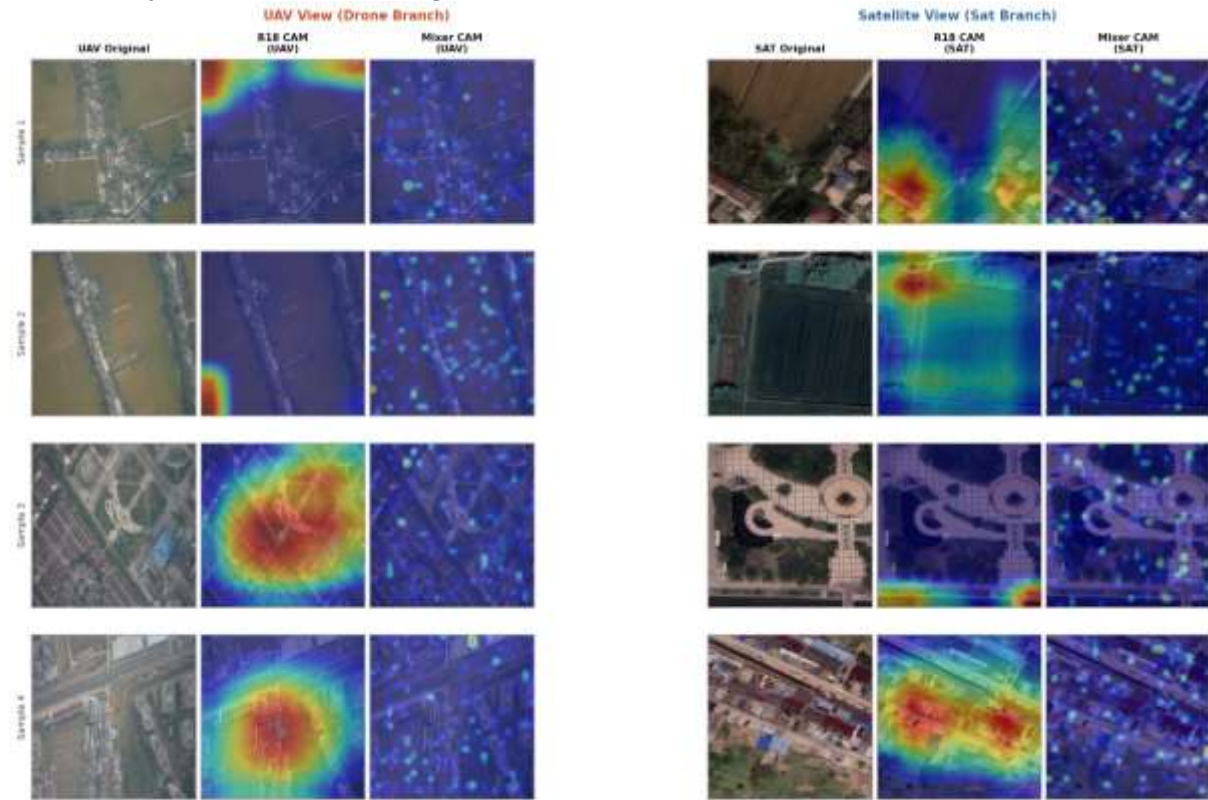


Figure 3. Grad-CAM attention heatmap visualization of ResNet-18 and MLP-Mixer. Warm colors indicate high-attention regions, cool colors indicate low-attention regions

Conclusion

This paper presents a systematic comparative study of dual-branch deep learning architectures for cross-view geo-localization, evaluating ResNet-18 and MLP-Mixer as backbones under a unified contrastive learning framework. Trained on UAV-satellite pairs with bidirectional InfoNCE loss and a shared 256-dimensional embedding space, we show that backbone architecture, loss function, weight-sharing, and pretraining each play distinct, measurable roles in retrieval performance.

Our experiments confirm MLP-Mixer [5] consistently outperforms ResNet-18 [16] across all metrics: R@1 of 74,09 % vs. 72,40 %, R@5 of 88,41 % vs. 87,50 %, and R@10 of 92,45 % vs. 90,49 %. This advantage stems from MLP-Mixer’s global token-mixing mechanism, which enables full-image spatial reasoning from the first layer critical for cross-view matching, where landmark layouts, not local textures, are the primary matching cue.

The ablation study reveals three key design principles: First, independent UAV satellite branches are essential, as shared weights cause a catastrophic 23,05 percentage point R@1 drop for ResNet-18, confirming the large domain gap between aerial and satellite imagery requires dedicated feature pathways. Second, InfoNCE loss significantly outperforms Triplet Loss for MLP-Mixer, as dense contrastive gradients from all batch pairs provide richer learning signals than single hard-negative mining [4]. Third, ImageNet pretraining is critical, especially for MLP-Mixer, whose R@1 collapses from 74,09 % to 20,05 % without initialization, a 54,04 percentage point drop, highlighting transfer learning’s necessity for inductive bias-free pure MLP architectures.

Qualitative analyses validate these findings: t-SNE visualizations [19] confirm tighter cross-view alignment with compact same-location embedding clusters, Grad-CAM maps [20] show MLP-Mixer focuses on semantic, viewpoint-invariant regions while ResNet-18 attends to view-sensitive low-level textures, and retrieval visualizations demonstrate higher-ranked correct matches for MLP-Mixer, especially in extreme viewpoint scale scenarios.

Despite these results, limitations remain: the 768-image dataset is small, performance on larger diverse datasets requires validation, the framework lacks GPS and temporal metadata for real-world deployment, and MLP-Mixer 59.9M parameter count is far higher than ResNet-18's 11.2M, posing constraints for resource-limited UAV systems.

Future work will focus on three directions: scaling to larger cross-view datasets to test generalizability, integrating lightweight attention/knowledge distillation to reduce MLP-Mixer computational cost while preserving performance, and extending to multi-modal fusion with GPS priors and flight trajectories for robust real-time UAV geo-localization.

In summary, this work provides a rigorous empirical foundation for cross-view geo-localization performance factors, establishing MLP-Mixer with bidirectional InfoNCE loss and independent dual branches as a strong baseline for future research.

References

- 1.°Shi Y., Liu L., Yu X., Li H. // *Advances in Neural Information Processing Systems (NeurIPS)*. 2019. Vol. 32. P. 10090–10100.
- 2.°Zhu S., Shah M., Chen C. // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2022. P. 1162–1171.
- 3.°Zheng Z., Wei Y., Yang Y. // *Proc. ACM Int. Conf. Multimedia (ACM MM)*. 2020. P. 1395–1403.
- 4.°Deuser F., Habel K., Oswald N. // *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*. 2023. P. 16847–16856.
- 5.°Tolstikhin I., Houlsby N., Kolesnikov A., Beyer L., Zhai X., Unterthiner T., Yung J., Steiner A., Keysers D., Uszkoreit J., Lucic M., Dosovitskiy A. // *Advances in Neural Information Processing Systems (NeurIPS)*. 2021. Vol. 34. P. 24261–24272.
- 6.°Workman S., Zhai M., Crandall D., Jacobs N. Wide-area image geolocalization with aerial reference imagery // *Proc. IEEE Int. Conf. Computer Vision (ICCV)*. 2015. P. 3961–3969.
- 7.°Zhu S., Shah M., Chen C. VIGOR: Cross-view image geo-localization beyond one-to-one retrieval // *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*. 2021. P. 3640–3649.
- 8.°Lowe D. G. Distinctive image features from scale-invariant keypoints // *Int. J. Comput. Vis.* 2004. Vol. 60. No. 2. P. 91–110.
- 9.°Dalal N., Triggs B. Histograms of oriented gradients for human detection // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2005. Vol. 1. P. 886–893.
- 10.°Arandjelović R., Gronat P., Torii A., Pajdla T., Sivic J. NetVLAD: CNN architecture for weakly supervised place recognition // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 5297–5307.
- 11.°Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. // *Proc. Int. Conf. Learning Representations (ICLR)*. 2021.
- 12.°Schroff F., Kalenichenko D., Philbin J. // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2015. P. 815–823.
- 13.°Chen T., Kornblith S., Norouzi M., Hinton G. // *Proc. Int. Conf. Machine Learning (ICML)*. 2020. Vol. 119. P. 1597–1607.
- 14.°van den Oord A., Li Y., Vinyals O. // *arXiv preprint arXiv:1807.03748*. 2018.
- 15.°Radenović F., Tolias G., Chum O. // *IEEE Trans. Pattern Anal. Mach. Intell.* 2019. Vol. 41. No. 7. P. 1655–1668.
- 16.°He K., Zhang X., Ren S., Sun J. // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 770–778.
- 17.°Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L. // *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*. 2009. P. 248–255.
- 18.°Loshchilov I., Hutter F. // *Proc. Int. Conf. Learning Representations (ICLR)*. 2019.
- 19.°van der Maaten L., Hinton G. // *J. Mach. Learn. Res.* 2008. Vol. 9. P. 2579–2605.
- 20.°Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. // *Proc. IEEE Int. Conf. Computer Vision (ICCV)*. 2017. P. 618–626.

СВЕДЕНИЯ ОБ АВТОРАХ

1. Бруттан Юлия Викторовна – к.т.н., доцент отделения информационно-коммуникационных технологий образовательного департамента Передовой инженерной школы гибридных технологий в станкостроении Союзного государства ПсковГУ
2. Ван Гуоянь – к.т.н., преподаватель Национального университета оборонных технологий, г. Чангша, Хунань, Китай
3. Ван Цань – студент кафедры инфокоммуникационных технологий БГУИР
4. Ван Цзяжоу – студент кафедры инфокоммуникационных технологий БГУИР
5. Ван Цикай – студент кафедры инфокоммуникационных технологий БГУИР
6. Ван Шуо – студент кафедры инфокоммуникационных технологий БГУИР
7. Васютин Андрей Дмитриевич – студент кафедры инфокоммуникационных технологий БГУИР
8. Венгеренко Вадим Владимирович – аспирант ОИПИ НАН Беларуси
9. Вишняков Владимир Анатольевич – д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
10. Воронов Александр Анатольевич – к.т.н., доц. лаб.222 ОИПИ НАН Беларуси
11. Го Цичэн – аспирант кафедры инфокоммуникационных технологий БГУИР
12. Горбуков Дмитрий Михайлович – к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР

- | | | |
|-----|-------------------------------|---|
| 13. | Воронович Дмитрий Юрьевич | – магистрант кафедры
инфокоммуникационных
технологий БГУИР |
| 14. | Жэнь Сюнь Хуань | – к.т.н., доцент кафедры
инфокоммуникационных
технологий БГУИР |
| 15. | Зорько Михаил Иванович | – ассистент кафедры
инфокоммуникационных
технологий БГУИР |
| 16. | Кацко Мария Олеговна | – магистрантка кафедры
инфокоммуникационных
технологий БГУИР |
| 17. | Кобяк Игорь Петрович | – к.т.н., доцент кафедры
электронных вычислительных
машин БГУИР |
| 18. | Курилович Андрей Владимирович | – старший преподаватель кафедры
инфокоммуникационных
технологий БГУИР |
| 19. | Левоненко Иван Игоревич | – магистрант факультета
информационной безопасности
БГУИР, инженер-исследователь
ОАО «Пеленг» |
| 20. | Линь Вей | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 21. | Литвинов Сергей Владимирович | – советник генерального
директора холдинга «АБВ», г.
Ростов-на-Дону, РФ |
| 22. | Ловецкий Михаил Юрьевич | – ассистент кафедры
инфокоммуникационных
технологий БГУИР |
| 23. | Ломако Александр Викторович | – кандидат технических наук,
доцент, доцент кафедры
информационных технологий
автоматизированных систем
БГУИР |
| 24. | Лукашевич Марина Михайловна | – к.т.н., доцент кафедры
информационных систем
управления БГУ |
| 25. | Лю Юй | – магистрант кафедры защиты
информации БГУИР |
| 26. | Ма Цзюнь | – к.т.н., ассистент кафедры
инфокоммуникационных
технологий БГУИР |

- | | | |
|-----|-------------------------------------|---|
| 27. | Макейчик Екатерина Геннадьевна | – старший преподаватель кафедры
инфокоммуникационных
технологий БГУИР |
| 28. | Максак Максим Юрьевич | – магистрант кафедры
инфокоммуникационных
технологий БГУИР |
| 29. | Масейчик Елена Алексеевна | – ассистент кафедры
инфокоммуникационных
технологий БГУИР |
| 30. | Петров Сергей Николаевич | – к.т.н., доцент кафедры защиты
информации БГУИР |
| 31. | Пилушко Александр Алексеевич | – к.т.н., доцент кафедры
инфокоммуникационных
технологий БГУИР |
| 32. | Пулко Татьяна Александровна | – к.т.н., доцент кафедры защиты
информации БГУИР |
| 33. | Робачевский Александр
Дмитриевич | – магистрант факультета
информационной безопасности
БГУИР |
| 34. | Смотрук Герман Сергеевич | – магистрант кафедры защиты
информации БГУИР, инженер-
системотехник ООО «СофтЛайн
Директ» |
| 35. | Соловьёв Алексей Николаевич | – научный сотрудник научно-
исследовательской лаборатории
прикладного анализа
инфокоммуникационных систем
кафедры телекоммуникационных
систем БГАС |
| 36. | Татур Михаил Михайлович | – д.т.н., профессор кафедры
электронных вычислительных
машин БГУИР |
| 37. | Тимофеев Александр Михайлович | – кандидат технических наук,
доцент, доцент кафедры защиты
информации БГУИР |
| 38. | Туманов Тимур Андреевич | – студент кафедры микро- и
наноэлектроники БГУИР |
| 39. | Фролов Григорий Артёмович | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 40. | Хаджинова Ксения Александровна | – студентка факультета
информационных технологий и
управления БГУИР |

- | | | |
|-----|-----------------------------------|---|
| 41. | Цветков Виктор Юрьевич | – заведующий кафедрой
инфокоммуникационных
технологий БГУИР, д.т.н.,
профессор |
| 42. | Чепикова Виолетта Викторовна | – старший преподаватель кафедры
инфокоммуникационных
технологий БГУИР |
| 43. | Чжан Бин | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 44. | Чжан Синьюй | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 45. | Чжан Чи | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 46. | Шао Синьюэ | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 47. | Шеметович Екатерина
Михайловна | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 48. | Шляхтун Дарья Витальевна | – учащаяся учреждения
образования «Национальный
детский технопарк» по
направлению «Информационная
безопасность» |
| 49. | Юй Цинхань | – студент кафедры
инфокоммуникационных
технологий БГУИР |
| 50. | Яковлев Юрий Сергеевич | – аспирант отделения
информационно-
коммуникационных технологий
ПсковГУ |
| 51. | Ян Цзяо | – студент кафедры
инфокоммуникационных
технологий БГУИР |