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, март – апрель 2023 г.)

Минск, 2023

Т31 Технологии передачи и обработки информации: материалы Международного научно-технического семинара (Минск, март – апрель 2023 г.) Technologies of information transmission and processing – Минск : БГУИР, 2023. – 176 с.
ISBN 978-985-488-834-7.

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

Корректор *В.В. Чепикова*
Ответственный за выпуск *В.Ю. Цветков*
Компьютерный дизайн и верстка *Е.Г. Макейчик*

Подписано в печать 08.04.2023. Формат 60×84 ⅛. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 10,46. Уч.-изд. л. 6,6. Тираж 50 экз. Заказ 760.

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

ISBN 978-985-488-834-7

СОДЕРЖАНИЕ

ОЦЕНКА КАЧЕСТВА СЖАТИЯ ДИНАМИЧЕСКОГО ДИАПАЗОНА ИНФРАКРАСНЫХ ИЗОБРАЖЕНИЙ
С.И. РУДИКОВ, В.Ю. ЦВЕТКОВ, А.П. ШКАДАРЕВИЧ 7
A HARD-DECISION ITERATIVE DECODING METHOD FOR 2D SEC-DED CODES
X.H. REN, Y.M. CHEN, V.K. KANAPELKA 15
NR LDPC CODES IN 5G MOBILE COMMUNICATION SYSTEM
F.Y. LIU, S.B. SALOMATIN 22
ФОРМИРОВАНИЕ КОМБИНИРОВАННЫХ АСМ-ИЗОБРАЖЕНИЙ НА ОСНОВЕ ВЗВЕШЕННОГО СЛОЖЕНИЯ ДВУХ КОМПОНЕНТ
М.Ю. ЛОВЕЦКИЙ, В.Ю. ЦВЕТКОВ, А.А. БОРИСКЕВИЧ, И.И. ЛЕВОНЕНКО, В.А. ЛАПИЦКАЯ, С.А. ЧИЖИК 27
HUMAN HEART RATE MONITORING BASED ON FACIAL VIDEO PROCESSING
N.V. BACH, I.A. BORISKIEVIC 34
VIDEO OBJECT DETECTION PROGRAM DESIGN UNDER TWO DIFFERENT CLIENT-SERVER ORGANIZATIONS
H. GAO, O.G. SHEVCHUK 39
ПОДХОД К ПОСТРОЕНИЮ ПОДСИСТЕМ УМНОГО ГОРОДА
В.А. ВИШНЯКОВ, В.А. ГРОМОВ, С.В. КУЧЕРОВ, С.А. СИДОРЕНКО, А.В. УСЕВИЧ 45
ANALYSIS AND SIMULATION BASED ON 5G CHANNEL CODING TECHNOLOGY
F.Y. LIU, S.B. SALOMATIN 50
СКЕЛЕТИРОВАНИЕ НИЗКОКОНТРАСТНЫХ ЗАШУМНЫХ СЕРЫХ ИЗОБРАЖЕНИЙ
Ц. МА, А.А. БОРИСКЕВИЧ 54
STRUCTURE AND COMPONENTS OF INTERNET OF THINGS NETWORK FOR IT PATIENT DIAGNOSTICS
U.A. VISHNYAKOU, H. TAO, Z. YIAN, W. HAORAN 61
DESIGN OF AUTOMATIC DISTRESS BRACELET FOR ELDERLY BASED ON SINGLE-CHIP MICROCOMPUTER
XU WEIXAUN, N.V. KHAJYNAVA 65
ПРОГРАММНАЯ МОДЕЛЬ ДВИЖЕНИЯ УЗЛОВ САМООРГАНИЗУЮЩЕЙСЯ СЕТИ В ТРЕХМЕРНОМ ПРОСТРАНСТВЕ
Т.В. ПОЛУЯН, С.Н. КАСАНИН 68
HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA AND CONVOLUTIONAL NEURAL NETWORK
Z. WAN, A.A. BARYSKIEVIC 72
УСИЛЕНИЕ ОПТИЧЕСКОЙ НЕСУЩЕЙ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРИЕМА В ВОЛОКОННО-ОПИЧЕСКИХ СИСТЕМАХ ПЕРЕДАЧИ
Я.В. РОЩУПКИН 78
SYSTEM DESIGN OVERVIEW OF HAMMING PRODUCT CODES IMPLEMENTATION ON FPGA
Y.M. CHEN, X.H. REN 84
ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ БИБЛИОТЕК РАСПОЗНАВАНИЯ ЛИЦ
С.Н. ПЕТРОВ, А.Д. МАТЮШЕНКО, Д.А. РУДЕНЯ 89
LAN INSTANT MESSAGE COMMUNICATION APPLICATION DESIGN BASED ON TCP
ZHANG BOWEN, ZHANG RONGLIANG, N.V. KHAJYNAVA 93
ПОТОКОВАЯ МОДЕЛЬ VPN С УЧЕТОМ ЗАДЕРЖКИ ПЕРЕДАЧИ ПАКЕТА В СЕТИ ЭЛЕКТРОСВЯЗИ СПЕЦИАЛЬНОГО НАЗНАЧЕНИЯ
С.С. ВРУБЛЕВСКИЙ, А.А. БЫСОВ 97
HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA CONVOLUTIONAL NERUAL NETWOEK AND LONG SHORT TIME MEMORY
Z.X. YANG, Z.Y. CHEN 102
АЛГОРИТМ КОДИРОВАНИЯ ПРОЦЕССА ТРАНСЛЯЦИИ БЕЛКОВ В КЛЕТКЕ
М.А. ПРОТЬКО, О.Ф. БОРИСЕНКО 108
RESEARCH ON WIRELESS AD HOC NETWORK TECHNOLOGY
Y. WANG, P. ZENG, Z.J. WEI 113
ЭКСПЕРИМЕНТАЛЬНЫЙ ПРОТОТИП ОТКРЫТОЙ СИСТЕМЫ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛЮДЕЙ ПРИ МНОГОКАМЕРНОМ ВИДЕОНАБЛЮДЕНИИ
С.А. ИГНАТЬЕВА, Н.А. ТОМАШЕВИЧ, А.А. ГОЛУБЕНОК, Р.П. БОГУШ 117

PHOTOPLETHYSMOGRAPHY AND ACCELEROMETER SENSORS SIGNALS FOR RECOGNIZING PHYSICAL ACTIVITY
S.S. WEI 123
ОБМЕН ИНФОРМАЦИЕЙ МЕЖДУ МОБИЛЬНЫМ ПРИЛОЖЕНИЕМ И МИКРОКОНТРОЛЛЕРОМ ЧЕРЕЗ ПРОТОКОЛ MQTT
А.В. ХАРЧЕНКО, В.С. ГАВРИЛЕНКО 128
DESIGN OF SPEECH RECOGNITION SYSTEM BASED ON ATTENTION MECHANISM
YALU GAO 132
SHORT MESSAGE SENDING PLATFORM BASED ON GSM MODEM
ZENG PENG, WEI ZIJIAN, WANG YING 135
АЛГОРИТМЫ СЛУЧАЙНОГО ПОИСКА В ОБУЧЕНИИ НЕЙРОННЫХ СЕТЕЙ
В.В. МАЦКЕВИЧ 139
INFORMATION SYSTEM DESIGN BASED ON MICROSERVICE ARCHITECTURE
HE RUNHAI, LI BOYI, ZHOU QUANHUA, ZHONG WU, ZHANG HENGRUI 145
RESEARCH ON TEXTURE IMAGE FEATURE EXTRACTION METHOD
J.K. CHEN, J.X. FU 149
DEVELOPMENT OF RECURRENT NEURAL NETWORKS
S.S. WEI 154
A HYBRID CLASSICATION ALGORITHM BASED ON SVM, ANN AND KNN FOR GESTURE RECOGNITION
Z.M. LIAO 159
HIGH DYNAMIC RANGE IMAGE PROCESSING TECHNOLOGY
J.X. FU, J.K. CHEN 163
ОПТИМИЗАЦИЯ ПРОЦЕССА МОНИТОРИНГА НЕИСПРАВНОСТЕЙ В КОММУТАТОРАХ СИСТЕМ ОПОВЕЩЕНИЯ
А.П. ТУРЛАЙ 167

CONTENTS

QUALITY EVALUATION OF DYNAMIC RANGE COMPRESSION OF INFRARED IMAGES
S.I. RUDIKOV, V.Yu. TSVIATKOU, A.P. SHKADAREVICH 7
A HARD-DECISION ITERATIVE DECODING METHOD FOR 2D SEC-DED CODES
X.H. REN, Y.M. CHEN, V.K. KANAPELKA 15
NR LDPC CODES IN 5G MOBILE COMMUNICATION SYSTEM
F.Y. LIU, S.B. SALOMATIN 22
FORMATION OF COMBINED AFM IMAGES BASED ON WEIGHTED ADDITION OF TWO COMPONENTS
M.YU. LAVETSKI, V.Yu. TSVIATKOU, A.A. BORISKEVICH, I.I. LIAVONENKA, V.A. LAPITSKAYA, S.A. CHIZHIK 27
HUMAN HEART RATE MONITORING BASED ON FACIAL VIDEO PROCESSING
N.V. BACH, I.A. BORISKIEVIC 34
VIDEO OBJECT DETECTION PROGRAM DESIGN UNDER TWO DIFFERENT CLIENT-SERVER ORGANIZATIONS
H. GAO, O.G. SHEVCHUK 39
AN APPROACH TO THE CONSTRUCTION OF SMART CITY SUBSYSTEMS
U.A. VISHNYAKOU, V.A. GROMOV, S.V. KUCHEROV, S.A. SIDORENKO, A.V. USEVICH 45
ANALYSIS AND SIMULATION BASED ON 5G CHANNEL CODING TECHNOLOGY
F.Y. LIU, S.B. SALOMATIN 50
SKELETING OF LOW-CONTRAST NOISY HALFTONE IMAGES
J MA, A.A. BORISKEVICH 54
STRUCTURE AND COMPONENTS OF INTERNET OF THINGS NETWORK FOR IT PATIENT DIAGNOSTICS
U.A. VISHNYAKOU, H. TAO, Z. YIAN, W. HAORAN 61
DESIGN OF AUTOMATIC DISTRESS BRACELET FOR ELDERLY BASED ON SINGLE-CHIP MICROCOMPUTER
XU WEIXAUN, N.V. KHAJYNAVA 65
MOTION SOFTWARE MODEL OF SELF-ORGANIZING NETWORK NODES IN THREE-DIMENSIONAL SPACE
T.V. POLUYAN, S.N. KASANIN 68
HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA AND CONVOLUTIONAL NEURAL NETWORK
Z. WAN, A.A. BARYSKIEVIC 72
OPTICAL CARRIER AMPLIFICATION TO IMPROVE THE RECEIVER EFFICIENCY IN FIBER-OPTIC COMMUNICATION SYSTEMS
Y.V. ROSHCHUPKIN 78
SYSTEM DESIGN OVERVIEW OF HAMMING PRODUCT CODES IMPLEMENTATION ON FPGA
Y.M. CHEN, X.H. REN 84
STUDY OF THE EFFECTIVENESS OF FACE RECOGNITION LIBRARIES
S.N. PETROV, A.D. MATSIUSHENKA, D.A. RUDENYA 89
LAN INSTANT MESSAGE COMMUNICATION APPLICATION DESIGN BASED ON TCP
ZHANG BOWEN, ZHANG RONGLIANG, N.V. KHAJYNAVA 93
HOSE MODEL OF VPN WITH DELAY IN A COMMUNICATION NETWORKS OF SPECIAL PURPOSE
S.S. VRUBLEVSKY, A.A. BYSOV 97
HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA CONVOLUTIONAL NERUAL NETWOEK AND LONG SHORT TIME MEMORY
Z.X. YANG, Z.Y. CHEN 102
ALGORITHM FOR ENCODING THE PROCESS OF PROTEIN TRANSLATION IN A CELL
M.A. PROTSKO, O.F. BORISENKO 108
RESEARCH ON WIRELESS AD HOC NETWORK TECHNOLOGY
Y. WANG, P. ZENG, Z.J. WEI 113
EXPERIMENTAL PROTOTYPE OF OPEN-WORLD PERSON RE-IDENTIFICATION SYSTEM IN MULTICAMERA VIDEO SURVEILLANCE
S.A. IHNATSYEVA, N.A. TOMASHEVICH, A.A. HALUBIONAK, R.P. BOHUSH 117

PHOTOPLETHYSMOGRAPHY AND ACCELEROMETER SENSORS SIGNALS FOR RECOGNIZING PHYSICAL ACTIVITY
S.S. WEI 123
EXCHANGE OF INFORMATION BETWEEN THE MOBILE APP AND THE MCU VIA MQTT PROTOCOL
A.V. HARCHENKO, V.S. GAVRILENKO 128
DESIGN OF SPEECH RECOGNITION SYSTEM BASED ON ATTENTION MECHANISM
YALU GAO 132
SHORT MESSAGE SENDING PLATFORM BASED ON GSM MODEM
ZENG PENG, WEI ZIJIAN, WANG YING 135
RANDOM SEARCH ALGORITHMS IN NEURAL NETWORKS TRAINING
V.V. MATSKEVICH 139
INFORMATION SYSTEM DESIGN BASED ON MICROSERVICE ARCHITECTURE
HE RUNHAI, LI BOYI, ZHOU QUANHUA, ZHONG WU, ZHANG HENGRUI 145
RESEARCH ON TEXTURE IMAGE FEATURE EXTRACTION METHOD
J.K. CHEN, J.X. FU 149
DEVELOPMENT OF RECURRENT NEURAL NETWORKS
S.S. WEI 154
A HYBRID CLASSICATION ALGORITHM BASED ON SVM, ANN AND KNN FOR GESTURE RECOGNITION
Z.M. LIAO 159
HIGH DYNAMIC RANGE IMAGE PROCESSING TECHNOLOGY
J.X. FU, J.K. CHEN 163
OPTIMIZATION OF THE PROCESS OF FAULT MONITORING IN THE SWITCHES OF WARNING SYSTEMS
A.P. TURLAI 167

УДК 621.391

ОЦЕНКА КАЧЕСТВА СЖАТИЯ ДИНАМИЧЕСКОГО ДИАПАЗОНА ИНФРАКРАСНЫХ ИЗОБРАЖЕНИЙ

С.И. РУДИКОВ[1], В.Ю. ЦВЕТКОВ[2], А.П. ШКАДАРЕВИЧ[1]

1 – Научно-технический центр «ЛЭМТ», Республика Беларусь
2 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Приведены результаты анализа чувствительности показателей качества тонового отображения к выбору алгоритмов сжатия динамического диапазона, их параметров и типов инфракрасных изображений. Показано, что интервальные показатели имеют большую чувствительность к условиям тонового отображения в сравнении с глобальными показателями.

Ключевые слова: сжатие динамического диапазона изображений, инфракрасные изображения, выравнивание гистограммы.

Введение

Для отображения и повышения качества воспроизведения многоразрядных инфракрасных изображений (ИК-изображений) на мониторе с меньшей разрядностью пикселей (как правило, 8 бит) широко используются алгоритмы преобразования динамического диапазона на основе выравнивания гистограммы (Histogram Equalization, HE) [1] и основанные на них блочные алгоритмы адаптивного выравнивания гистограммы (Adaptive Histogram Equalization, AHE) [2]. Алгоритм AHE обеспечивает достаточно высокое качество изображений после преобразования, но не позволяет управлять формой интегральной функции распределения, что приводит к чрезмерной контрастности для некоторых типов изображений.

В алгоритме адаптивного выравнивания гистограммы с ограничением контрастности (Contrast Limited Adaptive Histogram Equalization, CLAHE) [3] данная особенность учитывается за счет регулируемого ограничения гистограммы. Это позволяет управлять контрастностью в сторону ее уменьшения по сравнению с AHE. Но при снижении контрастности в алгоритме CLAHE ухудшаются и другие показатели качества изображения (среднее значение, средний градиент, энтропия, число деталей). Данный недостаток свойственен многим модификациям алгоритма AHE [4–6]. Причина заключается в выравнивании гистограммы, существенно ослабляющем результаты любой предкоррекции изображения. Поэтому эффективное управление характеристиками изображения при преобразовании динамического диапазона возможно только после выравнивания гистограммы. При этом для коррекции необходимо обеспечить достаточно широкий динамический диапазон изображения.

В [7] предложен алгоритм HECS (Histogram Equalization, Compression and Stretching, HECS), который превосходит алгоритм AHE по контрасту за счет за счет обрезки краев, растяжения центральной части, растяжения (сжатия) и наложения обрезанных краев глобальной гистограммы. Недостатком алгоритма HECS является ухудшение контраста по краям динамического диапазона, что приводит к заметным артефактам на преобразованном ИК-изображении и дополнительной неравномерности его глобальной гистограммы.

В [8] предложен модифицированный алгоритм HECSm адаптивной эквализации, растяжения и сжатия гистограммы, основанный на инверсии ее обрезанных краев. Инверсия краев позволяет сохранить корреляцию значений и контраст для большей части смежных пикселей изображения. При уменьшении динамического диапазона ландшафтных ИК-изображений предложенный алгоритм повышает блочный контраст по сравнению с

алгоритмом адаптивной эквализации гистограммы. Особенностью алгоритма является возможность растяжения одной части (левой или правой) за счет сжатия другой в пределах динамического диапазона. Это позволяет повысить разрешение в соответствующих интервалах гистограммы и локальный контраст в темных или светлых областях изображения. Для управления растяжением и сжатием краев гистограммы в алгоритме HECSm предусмотрен параметр A асимметрии гистограммы, вычисляемый как отношение ширины левого интервала гистограммы после преобразования к ширине динамического диапазона изображения.

Целью работы является определение показателей качества тонового отображения, имеющих наибольшую чувствительность к алгоритмам преобразования, их параметрам и типам ИК-изображений.

Показатели качества тонового отображения

Качество алгоритмов тонового отображения определяется качеством преобразованных с их помощью изображений (безэталонные показатели качества) и схожестью этих изображений с исходными изображениями (эталонные показатели качества). Благодаря простоте вычисления широко используются безэталонные показатели, позволяющие оценить контраст (стандартное отклонение D_{ST} и средний градиент G_{A}), энтропию E_{I}, количество локальных экстремумов N_{LE}, статистическую естественность N_{S} [9]. Для оценки качества тонового отображения на основе эталона в случае ИК-изображений с широким динамическим диапазоном часто используется показатели структурной точности F_{S} [10] и качества тональной карты I_{TMQ} [11]. Эти показатели не учитывают ряд характеристик сжатия динамического диапазона, связанных, например, с линейностью и последовательностью передачи тонов, потерей различения соседних пикселей после преобразования и равномерностью использования динамического диапазона, ростом неоднозначности тонового отображения из-за различий передаточных характеристик блоков изображения при использовании для преобразования блочных алгоритмов. Кроме того, большинство известных показателей вычисляются для всего динамического диапазона. Однако, в ряде случаев необходимы интервальные показатели, позволяющие оценить качество тонового отображения в определённой части динамического диапазона преобразованного ИК-изображения. Данные недостатки приводят к низкой точности и неоднозначности оценки качества сжатия динамического диапазона ИК-изображений. В [12] предложены интервальные показатели качества сжатия динамического диапазона ИК-изображений, позволяющие оценить:

– потенциальную различающую способность P_{D} на выбранном интервале динамического диапазона преобразованного изображения;

– потери E_{D} различения соседних пикселей на выбранном интервале динамического диапазона преобразованного изображения, обусловленные тоновым отображением;

– величину E_{MS} нелинейных искажений сжатия динамического диапазона на выбранном интервале динамического диапазона преобразованного изображения относительно линейно преобразованного изображения;

– равномерность U_{H} использования динамического диапазона на выбранном интервале динамического диапазона преобразованного изображения относительно базового интервала;

– неоднозначность L_{DH} тонового отображения, обусловленную различиями передаточных характеристик блоков в интервале динамического диапазона преобразованного изображения, соответствующего интервалу прореженного динамического диапазона исходного изображения;

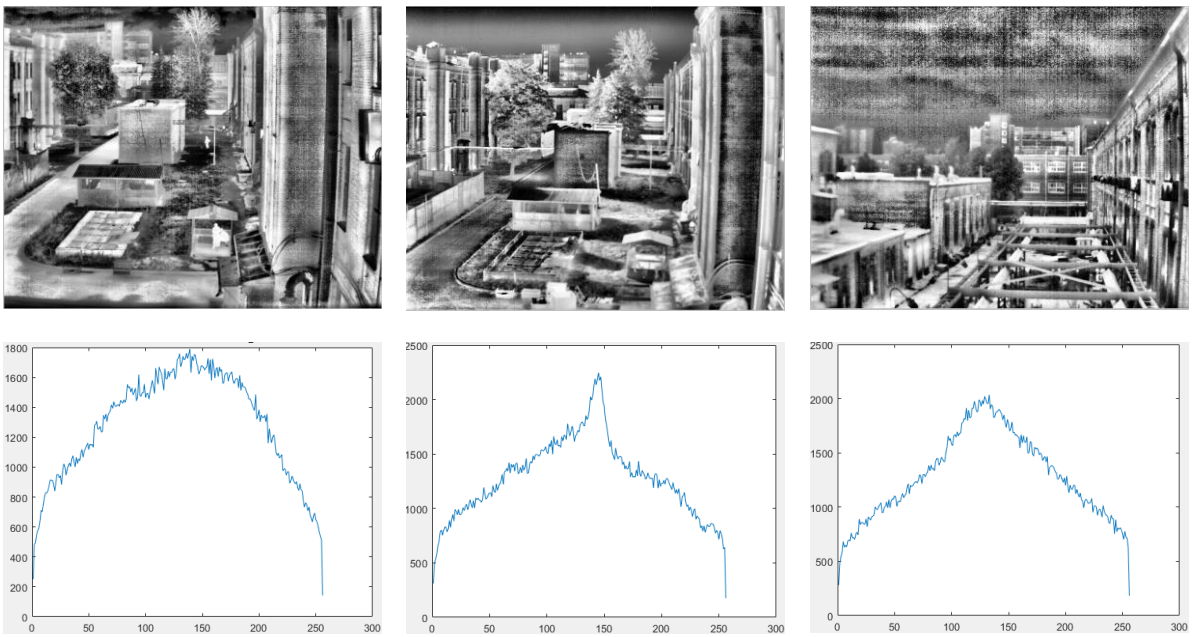
– величину L_{DL} нелинейных искажений, связанных с неоднозначностью тонового отображения, в интервале динамического диапазона преобразованного изображения, соответствующего интервалу прореженного динамического диапазона исходного изображения.

Чем меньше значения P_{D} и E_{D}, тем выше различающая способность и меньше потери различения соседних пикселей преобразованного изображения. Чем меньше значение E_{MS}, тем ближе передаточные характеристики блоков (или всего изображения) к линейным. Чем ближе к единице значение U_{H}, тем более равномерным является распределение яркостей на выбранном интервале относительного базового интервала и тем ближе тоновое отображение к линейному при

равновероятных значениях пикселей. Чем ближе к единице значение L_{DH}, тем меньше неоднозначность тонового отображения. Чем меньше значение L_{DL}, тем меньше нелинейные искажения из-за неоднозначности тонового отображения.

Оценка качества тонового отображения

На рис. 1 приведены примеры ИК-изображений трех типов, отличающихся формами гистограмм яркости после адаптивной эквализации. Для изображений этих трех типов в табл. 1 – 6 приведены средние значения отношений и разностей показателей качества для алгоритма HECSm при различных значениях асимметрии *A* и алгоритмов HE, AHE, HECS. Жирным шрифтом в таблицах отмечены максимальные значения отношений и разностей. Табл. 1, 3, 5 содержат значения глобальных параметров, определяемых в пределах всего динамического диапазона изображения. Табл. 2, 4, 6 содержат значения интервальных показателей, вычисляемых для перекрывающихся левого (L), центрального (C), правого (R) интервалов гистограммы и всего динамического диапазона изображения (интервал T). В алгоритмах AHE, HECS и HECSm использованы блоки 32×32 пикселя.



а *б* *в*

Рис. 1. Примеры ИК-изображений трех типов и их гистограмм после адаптивной эквализации: *а* – тип 1; *б* – тип 2; *в* – тип 3

Табл. 1. **Значения отношений глобальных показателей качества для изображений типа 1**

Показатели	Значения отношений показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS															
	HE					AHE					HECS					
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
N_S	**7,813**	7,348	6,844	6,542	6,235	**1,037**	0,975	0,909	0,868	0,828	**0,934**	0,879	0,819	0,783	0,746
F_S	0,933	0,961	**0,964**	0,960	0,942	0,963	0,992	**0,996**	0,991	0,973	1,001	1,031	**1,035**	1,030	1,012
I_{TMQ}	**1,123**	**1,123**	1,115	1,109	1,098	0,994	**0,995**	0,988	0,982	0,972	0,991	**0,992**	0,985	0,979	0,970
D_{ST}	1,890	1,882	1,860	1,871	**1,906**	0,961	0,957	0,946	0,952	**0,970**	0,908	0,904	0,893	0,899	**0,916**
G_A	**3,151**	3,082	3,023	3,036	3,092	**0,994**	0,972	0,954	0,957	0,975	**0,874**	0,855	0,838	0,842	0,857
E_I	1,209	**1,216**	1,215	1,212	1,203	0,983	**0,989**	0,988	0,986	0,978	0,980	**0,986**	0,985	0,983	0,976
N_{LE}	**1,169**	1,151	1,145	1,150	1,166	**1,047**	1,031	1,026	1,031	1,045	**0,955**	0,941	0,936	0,940	0,953

Использование классификации изображений на типы по форме гистограммы после адаптивной эквализации обусловлено особенностями формирования естественных

ИК-изображений, имеющих широкий динамический диапазон. Информация о распределении яркости на гистограммах таких изображений концентрируется в одном относительно узком интервале, что не позволяет различать особенности локального распределения в этом интервале. После эквализации гистограммы этот интервал растягивается практически на весь динамический диапазон. Причем, если алгоритм HE, не учитывающий локальные особенности изображения, обеспечивает близкое к равномерному распределение вероятностей значений пикселей в пределах динамического диапазона, то алгоритм AHE адаптивной эквализации, обрабатывающий изображение перекрывающимися блоками, частично сохраняет особенности распределения вероятностей исходных значений пикселей после преобразования.

Табл. 2. **Значения разностей интервальных показателей качества для изображений типа 1**

Показатели	Значения разностей показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS														
	HE					AHE					HECS				
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
Интервал L															
P_D	-78,9	12,5	60,6	92,2	**123,8**	-124,1	-32,7	15,3	47,0	**78,6**	-109,5	-18,1	29,9	61,6	**93,2**
E_D	-2,2	2,7	4,6	5,7	**6,7**	-4,5	0,3	2,3	3,3	**4,4**	-3,05	1,77	3,70	4,78	**5,83**
E_{MS}	-5,6	-1,4	2,0	4,3	**6,4**	-5,2	-0,9	2,5	4,7	**6,9**	-7,75	-3,49	-0,09	2,15	**4,30**
U_H	**-0,06**	-0,30	-0,44	-0,51	-0,55	-0,08	**-0,04**	-0,05	-0,06	-0,06	**0,05**	-0,19	-0,33	-0,40	-0,45
L_{DH}	**-86,7**	-107,7	-108,8	-105,1	-98,4	**21,5**	0,5	-0,5	3,1	9,8	14,31	-6,71	-7,77	-4,12	**2,63**
L_{DL}	**-0,31**	-0,56	-0,69	-0,78	-0,86	**0,26**	0,01	-0,12	-0,21	-0,29	0,19	**-0,06**	-0,19	-0,29	-0,36
Интервал C															
P_D	-73,9	-80,8	-87,1	-86,3	**-69,5**	-1,7	-8,7	-15,0	-14,1	**2,7**	-18,9	-25,8	-32,1	-31,3	**-14,5**
E_D	-21,9	-22,8	**-23,3**	-22,6	-20,5	-0,1	-1,0	-1,5	-0,8	**1,3**	-2,06	-3,01	-3,44	-2,76	**-0,65**
E_{MS}	-3,3	-0,3	1,3	2,3	**3,2**	-2,4	0,7	2,2	3,2	**4,1**	-4,39	-1,32	0,22	1,22	**2,13**
U_H	-0,32	-0,66	-0,74	-0,57	**-0,20**	0,02	-0,33	-0,40	-0,24	**0,13**	-0,16	-0,50	-0,58	-0,41	**-0,04**
L_{DH}	-89,6	-83,7	-73,0	-64,2	**-52,5**	-1,3	4,6	15,3	24,1	**35,8**	-23,11	-17,19	-6,51	2,34	**14,04**
L_{DL}	-0,38	-0,33	-0,29	-0,26	**-0,24**	0,00	0,05	0,09	0,12	**0,14**	-0,05	-0,01	0,04	0,06	**0,09**
Интервал R															
P_D	**129,0**	84,6	33,7	-32,5	-134,5	**107,0**	62,6	11,7	-54,4	-156,5	**125,9**	81,5	30,6	-35,6	-137,6
E_D	25,3	23,9	22,0	18,4	12,4	**4,7**	3,4	1,4	-2,1	-8,1	**7,26**	5,91	3,97	0,39	-5,59
E_{MS}	-0,4	**0,2**	**0,2**	0,0	-0,3	-0,5	**0,0**	**0,0**	-0,2	-0,5	-0,22	0,30	**0,32**	0,17	-0,18
U_H	-0,55	-0,48	-0,39	**-0,20**	-0,22	-0,33	-0,25	-0,16	**0,03**	0,01	-0,43	-0,36	-0,27	**-0,08**	-0,10
L_{DH}	-28,3	-25,4	-23,6	-21,6	**-18,4**	-2,7	0,2	2,0	4,0	**7,1**	-0,05	2,79	4,57	6,57	**9,77**
L_{DL}	-0,37	-0,27	-0,20	-0,14	**-0,07**	-0,17	-0,08	-0,01	0,06	**0,12**	-0,14	-0,04	0,02	0,09	**0,13**
Интервал T															
E_{MS}	-3,9	0,1	2,4	3,9	**5,0**	-4,9	-0,9	1,4	2,9	**4,0**	-4,40	-0,46	1,87	3,36	**4,50**
L_{DH}	-67,1	-68,0	-67,8	-67,3	**-66,2**	2,2	1,2	1,5	1,9	**3,1**	5,00	4,03	4,29	4,73	**5,85**
L_{DL}	-0,35	-0,35	-0,35	-0,35	**-0,34**	0,01	0,01	0,01	0,01	**0,02**	0,10	0,10	0,11	0,11	**0,12**

Табл. 3. **Значения отношений глобальных показателей качества для изображений типа 2**

Показатели	Значения отношений показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS														
	HE					AHE					HECS				
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
N_S	**16,82**	15,99	14,84	13,97	13,15	1,097	1,042	0,967	0,910	0,858	**0,931**	0,884	0,821	0,773	0,728
F_S	0,866	0,904	**0,909**	0,904	0,883	0,945	0,986	**0,992**	0,986	0,963	0,985	1,027	**1,034**	1,028	1,003
I_{TMQ}	1,120	**1,125**	1,118	1,109	1,095	0,996	**1,001**	0,994	0,986	0,974	0,988	**0,992**	0,986	0,978	0,966
D_{ST}	2,433	2,443	2,419	2,431	**2,469**	0,993	0,996	0,987	0,992	**1,007**	0,903	0,906	0,897	0,902	**0,916**
G_A	**4,221**	4,131	4,041	4,042	4,101	1,038	**1,014**	0,992	0,993	1,007	**0,878**	0,858	0,839	0,840	0,852
E_I	1,260	**1,269**	1,267	1,263	1,251	0,985	**0,992**	0,990	0,987	0,978	0,975	**0,982**	0,981	0,977	0,968
N_{LE}	**1,195**	1,177	1,170	1,175	1,188	1,050	1,034	1,029	1,032	**1,044**	**0,955**	0,941	0,936	0,940	0,950

Табл. 4. **Значения разностей интервальных показателей качества для изображений типа 2**

Показатели	Значения разностей показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS														
	HE					AHE					HECS				
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
Интервал L															
P_D	-107,5	14,0	77,2	119,9	**185,5**	-169,3	-47,8	15,4	58,1	**123,7**	-131,6	-10,1	53,1	95,8	**161,4**
E_D	-23,05	-17,05	-14,87	-13,37	**-11,72**	-6,50	-0,49	1,68	3,18	**4,83**	-4,78	1,22	3,39	4,89	**6,54**
E_{MS}	-3,67	0,25	3,59	**6,05**	3,75	-6,16	-2,24	1,10	**3,55**	1,26	-9,09	-5,17	-1,83	**0,63**	-1,67
U_H	**-0,08**	-0,22	-0,40	-0,49	-0,53	0,21	**0,06**	-0,11	-0,20	-0,24	**0,00**	-0,14	-0,31	-0,40	-0,44
L_{DH}	**-82,84**	-92,64	-98,84	-96,60	-87,45	11,28	1,49	-4,71	-2,48	**6,68**	**15,60**	5,81	-0,39	1,84	11,00
L_{DL}	**-0,29**	-0,52	-0,73	-0,88	-1,11	0,31	**0,08**	-0,13	-0,28	-0,51	**0,24**	0,00	-0,20	-0,35	-0,58
Интервал C															
P_D	-100,0	-106,7	-115,1	-112,3	**-21,8**	7,0	0,4	-8,0	-5,2	**85,2**	-40,4	-47,0	-55,4	-52,6	**37,8**
E_D	-25,08	-26,60	-26,95	-26,45	**-21,37**	0,41	-1,12	-1,46	-0,96	**4,45**	-1,43	-2,96	-3,30	-2,80	**2,28**
E_{MS}	-2,70	0,48	2,11	**3,16**	2,64	-3,24	-0,06	1,57	**2,62**	2,10	-5,20	-2,02	-0,39	**0,67**	0,15
U_H	-0,28	-0,64	-0,75	-0,57	**-0,13**	0,19	-0,17	-0,28	-0,10	**0,34**	-0,11	-0,47	-0,58	-0,40	**0,04**
L_{DH}	-76,05	-74,17	-67,09	-59,02	**-57,61**	-2,13	-0,25	6,83	14,91	**16,31**	-18,91	-17,03	-9,95	-1,88	**-0,47**
L_{DL}	-0,16	-0,14	-0,13	**-0,11**	-0,15	-0,02	-0,01	0,01	**0,02**	-0,04	-0,03	0,04	0,07	**0,08**	-0,02
Интервал R															
P_D	**179,2**	122,2	52,6	-52,7	-215,4	**136,6**	79,6	9,9	-95,3	-258,1	**164,1**	107,1	37,5	-67,8	-230,5
E_D	**38,40**	36,48	34,45	29,64	19,42	**5,31**	3,38	1,36	-3,45	-13,67	**7,97**	6,04	4,02	-0,79	-11,01
E_{MS}	-0,33	**0,01**	-0,13	-0,43	-0,65	-0,48	**-0,14**	-0,28	-0,58	-0,80	0,22	**0,56**	0,42	0,12	-0,10
U_H	-0,54	-0,46	-0,36	**-0,14**	-0,50	-0,28	-0,20	-0,10	**0,12**	-0,24	-0,41	-0,33	-0,24	**-0,01**	-0,38
L_{DH}	-23,22	-17,80	-13,83	**-11,91**	-12,01	-2,97	2,45	6,43	**8,34**	8,24	-10,22	-4,80	-0,83	**1,08**	0,98
L_{DL}	-0,29	-0,21	-0,13	-0,07	**-0,04**	-0,13	-0,04	0,02	0,06	**0,07**	-0,18	-0,10	-0,02	0,03	**0,07**
Интервал T															
E_{MS}	-4,44	-0,64	1,77	3,34	**4,30**	-5,47	-1,68	0,73	2,31	**3,27**	-4,80	-1,01	1,40	2,98	**3,94**
L_{DH}	-61,21	-61,84	-61,46	-60,86	**-59,46**	0,91	0,27	0,65	1,25	**2,65**	4,53	3,90	4,28	4,88	**6,28**
L_{DL}	-0,37	-0,37	-0,37	**-0,36**	**-0,36**	0,01	0,00	0,00	0,01	**0,02**	0,03	0,02	0,02	0,03	**0,04**

Табл. 5. **Значения отношений глобальных показателей качества для изображений типа 3**

Показатели	Значения отношений показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS														
	HE					AHE					HECS				
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
N_S	**18,406**	17,478	16,284	15,344	15,167	**1,081**	1,026	0,953	0,900	0,847	**0,928**	0,881	0,819	0,773	0,727
F_S	0,964	0,993	**0,994**	0,985	0,964	0,959	0,990	**0,993**	0,986	0,960	0,999	1,032	**1,035**	1,027	1,001
I_{TMQ}	1,154	**1,155**	1,147	1,137	1,127	0,998	**1,000**	0,993	0,985	0,972	0,990	**0,993**	0,985	0,977	0,964
D_{ST}	**2,221**	2,206	2,178	2,184	2,209	**0,992**	0,986	0,971	0,973	0,985	**0,902**	0,896	0,883	0,884	0,895
G_A	**4,341**	4,230	4,140	4,144	4,213	**1,031**	1,003	0,979	0,978	0,992	**0,867**	0,843	0,823	0,823	0,834
E_I	1,275	**1,281**	1,278	1,272	1,264	0,986	**0,991**	0,989	0,985	0,977	0,976	**0,981**	0,979	0,976	0,967
N_{LE}	**1,209**	1,191	1,186	1,191	1,205	**1,047**	1,031	1,025	1,029	1,041	**0,960**	0,945	0,940	0,943	0,954

Из табл. 1, 3, 5 следует, что статистическая естественность N_S достаточно чувствительна к изменению асимметрии алгоритма HECSm и выбору алгоритма между HE и AHE (HECS, HECSm), но слабо зависит от типа изображения и выбора алгоритма между AHE, HECS и HECSm. Структурная точность F_S достаточно чувствительна к изменению асимметрии алгоритма HECSm, выбору алгоритма между HE и AHE (HECS, HECSm) для некоторых типов изображений и выбору алгоритма между AHE, HECS и HECSm. Качество тональной карты I_{TMQ} достаточно чувствительно к изменению асимметрии алгоритма HECSm и выбору алгоритма между HE и AHE (HECS, HECSm).

Табл. 6. **Значения разностей интервальных показателей качества для изображений типа 3**

Показатели	Значения разностей показателей для алгоритма HECSm при различных *A* и алгоритмов HE, AHE, HECS														
	HE					AHE					HECS				
	A=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5	*A*=1/5	*A*=1/3	*A*=1/2	*A*=2/3	*A*=4/5
Интервал L															
P_D	-85,46	48,79	119,41	164,37	**202,28**	-203,12	-63,82	22,75	52,61	**123,42**	-152,06	-12,60	66,31	103,79	**174,61**
E_D	-20,03	-10,61	-5,83	**-4,17**	-5,04	-5,08	4,53	11,76	11,75	**12,76**	-2,52	7,12	13,99	14,30	**15,14**
E_{MS}	-4,30	-0,16	3,23	5,50	**3,44**	-6,44	-2,21	-1,50	**3,75**	2,56	-9,93	-5,70	-4,44	**0,27**	-0,94
U_H	**-0,12**	-0,29	-0,45	-0,53	-0,52	**0,23**	0,05	-0,12	-0,20	-0,24	**0,02**	-0,16	-0,33	-0,41	-0,45
L_{DH}	-73,67	**-70,46**	-70,61	-72,24	-75,43	-2,65	**1,23**	1,08	-1,91	-6,35	14,69	**20,52**	20,30	15,89	9,54
L_{DL}	**-0,39**	-0,54	-0,68	-0,83	-1,07	**0,22**	0,08	-0,12	-0,20	-0,54	**0,22**	0,08	-0,13	-0,20	-0,54
Интервал C															
P_D	-124,7	-131,48	-138,70	-130,10	**-42,68**	-0,12	-6,80	-31,71	-11,10	**86,48**	-54,78	-61,17	-79,35	-65,45	**32,49**
E_D	-21,84	-25,97	-26,10	-25,68	**-19,67**	-3,46	-5,89	-7,88	-5,80	**3,11**	-7,38	-9,80	-11,37	-9,79	**-1,02**
E_{MS}	-2,51	0,81	2,56	**3,65**	2,97	-3,54	-0,06	0,94	**2,88**	2,84	-6,17	-2,69	-1,30	**0,25**	0,20
U_H	-0,34	-0,67	-0,74	-0,52	**-0,17**	0,11	-0,15	-0,23	-0,09	**0,23**	-0,14	-0,50	-0,60	-0,42	**0,01**
L_{DH}	-73,94	-70,71	-65,63	**-63,41**	-64,05	-2,38	0,37	5,46	8,15	**8,56**	-7,96	-5,25	-0,14	2,54	**2,94**
L_{DL}	-0,43	-0,38	-0,34	**-0,33**	-0,41	-0,03	0,02	0,00	**0,07**	-0,01	-0,02	0,03	0,01	**0,09**	0,01
Интервал R															
P_D	176,19	115,87	34,81	-85,38	-217,58	170,19	110,63	36,73	-74,86	-264,25	193,43	134,15	55,65	-51,42	-241,27
E_D	41,94	40,20	35,98	23,93	12,74	9,86	8,24	5,64	-3,08	-25,07	17,66	16,06	12,32	4,90	-16,85
E_{MS}	-0,58	-0,03	0,01	-0,17	-0,30	-0,69	-0,10	-0,12	-0,20	-0,31	-0,43	0,16	0,09	0,05	-0,06
U_H	-0,57	-0,50	-0,38	-0,22	-0,51	-0,30	-0,23	-0,13	0,09	-0,21	-0,43	-0,35	-0,25	-0,04	-0,34
L_{DH}	-38,22	-25,32	-21,58	-20,46	-24,11	-14,52	-0,66	3,19	4,60	1,26	-18,56	-4,73	-0,70	0,66	-2,84
L_{DL}	-0,53	-0,30	-0,19	-0,13	-0,13	-0,34	-0,10	-0,01	0,08	0,10	-0,39	-0,15	-0,05	0,03	0,05
Интервал T															
E_{MS}	-4,22	-0,27	2,26	3,93	4,01	-5,90	-1,78	-0,38	2,49	3,68	-5,35	-1,23	0,09	3,04	4,22
L_{DH}	-59,53	-59,89	-59,55	-58,96	-57,87	0,67	0,25	0,59	1,09	2,27	4,09	3,67	4,01	4,51	5,70
L_{DL}	-0,42	-0,42	-0,42	-0,42	-0,41	0,01	0,00	-0,05	0,01	0,02	0,03	0,03	-0,03	0,03	0,04

Стандартное отклонение D_{ST}, средний градиент G_A и количество локальных экстремумов N_{LE} достаточно чувствительны к изменению асимметрии алгоритма HECSm, выбору алгоритма эквализации и выбору типа изображения при использовании алгоритма HE. Энтропия E_I достаточно чувствительна к выбору алгоритма между HE и AHE (HECS, HECSm), выбору алгоритма между AHE, HECS и HECSm для некоторых типов изображений, выбору типа изображения при использовании алгоритма HE. Результаты анализа глобальных показателей приведены в табл. 7.

Табл. 7. **Чувствительность глобальных показателей в различных условиях**

Показатели	Высокая (+) или низкая (–) чувствительность глобальных показателей в зависимости от условий			
	Изменение значения асимметрии алгоритма HECSm	Использование алгоритмов HE или AHE (HECS, HECSm)	Использование алгоритмов AHE, HECS или HECSm	Использование изображений различного типа
N_S	+	+	–	–
F_S	+	+/–	+	–
I_{TMQ}	+	+	+	–
D_{ST}	+	+	+	+/–
G_A	+	+	+	+/–
E_I	–	+	+/–	+/–
N_{LE}	+	+	+	+/–

Из табл. 2, 4, 6 следует, что потенциальная различающая способность P_D, потери E_D различения соседних пикселей, величина E_{MS} нелинейных искажений сжатия динамического диапазона равномерность U_H использования динамического диапазона, неоднозначность L_{DH} тонового отображения, величину L_{DL} нелинейных искажений достаточно чувствительны к изменению асимметрии алгоритма HECSm, выбору алгоритма эквализации, типа изображения и интервала оценки (табл. 8).

Табл. 8. **Чувствительность интервальных показателей в различных условиях**

Показа-тели	Высокая (+) или низкая (–) чувствительность глобальных показателей в зависимости от условий				
	Изменение значения асимметрии алгоритма HECSm	Использование алгоритмов HE или AHE (HECS, HECSm)	Использование алгоритмов AHE, HECS или HECSm	Использование изображений различного типа	Выбор интервала оценки
P_D	+	+	+	+	+
E_D	+	+	+	+	+
E_{MS}	+	+	+	+	+
U_H	+	+	+	+	+
L_{DH}	+	+	+	+	+
L_{DL}	+	+	+	+	+

Из табл. 7, 8 следует, что интервальные показатели превосходят глобальные по чувствительности и могут эффективно использоваться для оценки качества тонового отображения. Из глобальных показателей качества тонового отображения наиболее эффективными являются стандартное отклонение D_{ST}, средний градиент G_A и количество локальных экстремумов N_{LE}, имеющие высокую чувствительность при относительно большом разнообразии условий преобразования.

Заключение

Произведена оценка чувствительности показателей качества сжатия динамического диапазона ИК-изображений к алгоритмам преобразования, их параметрам и особенностям гистограмм яркости пикселей. Рассмотрены глобальные и интервальные показатели качества. Показано, что интервальные показатели превосходят глобальные по чувствительности.

QUALITY EVALUATION OF DYNAMIC RANGE COMPRESSION OF INFRARED IMAGES

S.I. RUDIKOV, V.Yu. TSVIATKOU, A.P. SHKADAREVICH

Abstract. The results of sensitivity analysis of tone mapping quality indicators to the choice of dynamic range compression algorithms, their parameters and types of infrared images were presented. It was shown that interval indicators are more sensitive to tone mapping conditions in comparison with global indicators.

Keywords: image dynamic range compression, infrared images, histogram equalization.

Список литературы

1. Nithyananda C.R., Ramachandra A.C., Preethi // 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). – Chennai, 2016. – P. 2512–2517.
2. Kim T.K., Paik J.K., Kang B.S. // IEEE Trans. Consum. Electron. – 1998. – Vol. 44, № 1. – P. 82–87.
3. Reza A.M. // Journal of VLSI Signal Process.-Syst. Signal Image Video Technol. – 2004. – Vol. 38, № 1. – P. 35–44.
4. Kim J.-Y., Kim L.-S., Hwang S.-H. // IEEE Transactions on Circuits and Systems for Video Technology. – 2001. – Vol. 11, № 4. – P. 475–484.

5. Huang S.-C., Yeh C.-H. // Engineering Applications of Artificial Intelligence. – 2013. – Vol. 26, № 5. – P. 1487–1492.
6. Al-Sammaraie M. F. // 10th International Conference on Computer Science and Education (ICCSE). – Cambridge, 2015. – P. 95–101.
7. Рудиков С.И., Цветков В.Ю., Шкадаревич А.П. // Весці Нацiянальнай акадэмii навук Беларусі. Серыя фізіка-матэматычных навук. 2021. Т. 66. № 4. С. 470–482.
8. Рудиков С.И., Цветков В.Ю., Шкадаревич А.П. // Технологии передачи и обработки информации: материалы международного научно-технического семинара, Минск, март-апрель 2022 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2022. – С. 5–11.
9. Mante V., Frazor R. A., Bonin V., Geisler W.S., Carandini M. // (2005) Independence of luminance and contrast in natural scenes and in the early visual system. Nat Neurosci, 8, 1690–1697.
10. Wang Z., Simoncelli E. P., Bovik A. C. // (2003) Multiscale structural similarity for image quality assessment. 37th Asilomar Conference on Signals, Systems & Computers. Pacific Grove, CA, USA, 2, 1398–1402.
11. Yeganeh H., Wang Z. // (2013) Objective Quality Assessment of Tone-Mapped Images. IEEE Transactions on Image Processing, 22(2), 657–667.
12. Рудиков С.И., Цветков В.Ю., Шкадаревич А.П. // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки, (11), 30-39.

UDC 621.391.(075.8)

A HARD-DECISION ITERATIVE DECODING METHOD FOR 2D SEC-DED CODES

X.H. REN, Y.M. CHEN, V.K. KANAPELKA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 13, 2023

Abstract. Product codes are well known to have very promising correction potential and the ability to deal with burst errors and can correct random errors in which the number of errors is above half the minimum distance. Two-dimensional single-error correcting and double-error detecting (2D SEC-DED) codes are types of product codes. Decoding methods for product codes can roughly be divided into hard-decision decoding and soft-decision decoding. The hard-decision decoding method, also called the iterative decoding method, which is easy to implement and has low computational complexity. However, the error correcting capability of the existing hard-decision iterative decoding methods for 2D SEC-DED is much less than half the minimum distance of the code. In this paper, we propose an improved hard-decision iterative decoding method for 2D SEC-DED codes to overcome this defect.

Keywords: hard-decision, 2D SEC-DED codes, iterative decoding.

Introduction

A 2D SEC-DED code based on two extended Hamming codes $C_1(n_1,k_1,4)$ and $C_2(n_2,k_2,4)$ can be constructed by performing the following two steps. Suppose that *k* bits of the message can be resized to a rectangular array with k_1 rows and k_2 columns. We first encode the message in the row direction by using the encoding rule of extended Hamming code C_2 and obtain a k_1 row and n_2 column code C_{mid}. Then, we encode C_{mid} in the column direction by using the encoding rule of the extended Hamming code C_1 and obtain the final two-dimensional code C_{2d} with parameters $(n_1n_2,k_1k_2,4)$. We know that the maximum correction capability of a 2D SEC-DED code based on extended Hamming codes is seven. Therefore, a proper iterative hard-decision decoding method for 2D SEC-DED codes should satisfy the following two requirements. First, it should be able to correct all error patterns in which the number of error bits is less than or equal to seven. Second, the iterative hard-decision decoding method should correct as many error patterns as possible in which the number of error bits is greater than seven [1–5].

Decoding methods for 2D SEC-DED codes

The former decoding methods usually adopt an iterative approach, as first introduced by Elias [1]. There are two major types of decoding methods for product codes: hard-decision decoding and soft-decision decoding. Iterative decoding methods are easy to implement and have low computational complexity [2–5]. In the past decade, many efforts have been made and many improved methods have been proposed to improve the performance of hard-decision iterative decoding [6–12]. In comparison, decoding methods based on the soft-decision approach perform better in terms of correctness capability than hard-decision decoding, but they also have high complexity and require extra information to indicate the reliability of each piece of input data. The maximum correction capability of iterative decoding methods is up to half the minimum distance of the code. However, the limitation of iterative decoding is that it is unable to correct certain special error patterns, named stall patterns, for which the weight is within half the minimum distance [11].

Two-dimensional single-error correcting and double-error detecting (2D SEC-DED) codes are types of product codes, which are widely adopted in many applications, and the corresponding encoding and decoding procedures are relatively simple. One SEC-DED code is the extended Hamming code, which can be obtained by adding one extra parity bit to the original Hamming code. Many distinct methods have been proposed in the past. However, the error correction capabilities of most of them are insufficient to correct error patterns with the number of errors up to half the minimum distance of the code. In our previous work [13], we designed an iterative decoding method for standard Hamming product codes that can correct all stall patterns with four errors and thus can correct all errors up to half the minimum distance. However, this method is not suitable for 2D SEC-DED codes because the component codes are different, and the minimum distance correspondingly increases from the original value of four to seven. To overcome this challenge, in this paper, we follow the idea of our previous work and design an improved hard-decision iterative decoding method for 2D SEC-DED codes based on extended Hamming codes; this approach can be used to correct errors up to half the minimum distance of the code.

The simplest iterative hard-decision decoding method is the two-step row-column method [3]. In the first step, the syndromes of all columns of the received code are computed in accordance with the decoding method corresponding to the encoding method, based on which the decoder locates all possible positions of single errors (correctable errors) and rectifies them in place. The decoder will not attempt to fix any double errors (uncorrectable errors). Then, the decoding result of the first step is passed to the second step. In the second step, a similar decoding operation is performed again but in the other direction. The two-step decoding method is very efficient and can correct many error patterns with the number of errors above half the minimum distance of the code. However, it fails to correct some stall patterns, such as 2-by-2 error patterns, since the decoder takes no action for double errors.

The proposed decoding method

The proposed method consists of two procedures: a preprocessing procedure and a decoding procedure.

1. Preprocessing procedure.

In the preprocessing procedure, in addition to the registers for the received code, four additional registers are required to record the error status: the row existing-error register (REER), the row double-error register (RDER), the column existing-error register (CEER) and the column double-error register (CDER). The *i*-th bit of the REER/CEER will be set to one when errors are detected in the *i*-th row/column based on the syndrome. Otherwise, this bit should be set to 0. Similarly, the *i*-th bit of the RDER/CDER will be set to one only when a double error is detected in the *i*-th row/column according to the syndrome. Otherwise, this bit should be set to 0.

The preprocessing procedure has two functions: determining the initial decoding direction and applying a pre-erasure process to reduce the number of errors when necessary. The initial direction of decoding is determined by comparing the estimated numbers of errors from rows ($\mathrm{RN}_{\mathrm{error}}$) with the estimated numbers of errors from columns ($\mathrm{CN}_{\mathrm{error}}$), which can be computed according to formulas (1) and (2), respectively. If $\mathrm{RN}_{\mathrm{error}}$ is greater than $\mathrm{CN}_{\mathrm{error}}$, then the initial decoding direction remains the row direction (the default direction); in contrast, if $\mathrm{RN}_{\mathrm{error}}$ is less than $\mathrm{CN}_{\mathrm{error}}$, then the initial decoding direction is changed to the column direction by transposing the received code (a flag will be set to indicate whether transposition is conducted; if the flag is true, then at the end of the decoding procedure, another transposition is conducted after the whole decoding process is complete). The reason for this is that we contend that decoding from the side from which more errors are estimated initially will introduce fewer errors during the decoding procedure.

$$\mathrm{RN}_{\mathrm{error}} = \sum_{i=1}^{n_2} \mathrm{REER}_i + \sum_{i=1}^{n_2} \mathrm{RDER}_i \tag{1}$$

$$CN_{error} = \sum_{i=1}^{n_1} CEER_i + \sum_{i=1}^{n_1} CDER_i \quad (2)$$

The pre-erasure process is implemented when the following three conditions are satisfied: RN_{error} is equal to CN_{error}; the numbers of instances of 1 in both the REER and CEER are equal; and the product of the numbers of instances of 1 in the REER and CEER is less than the sum of RN_{error} and CN_{error}. The positions for erasure are determined by the values of 1 in the REER and in CEER. The idea behind this is intuitive: performing the erasure process on a small region in which the number of error bits is greater than the number of correct bits can reduce the number of error bits. The flow chart of the preprocessing procedure is presented in Figure 1.

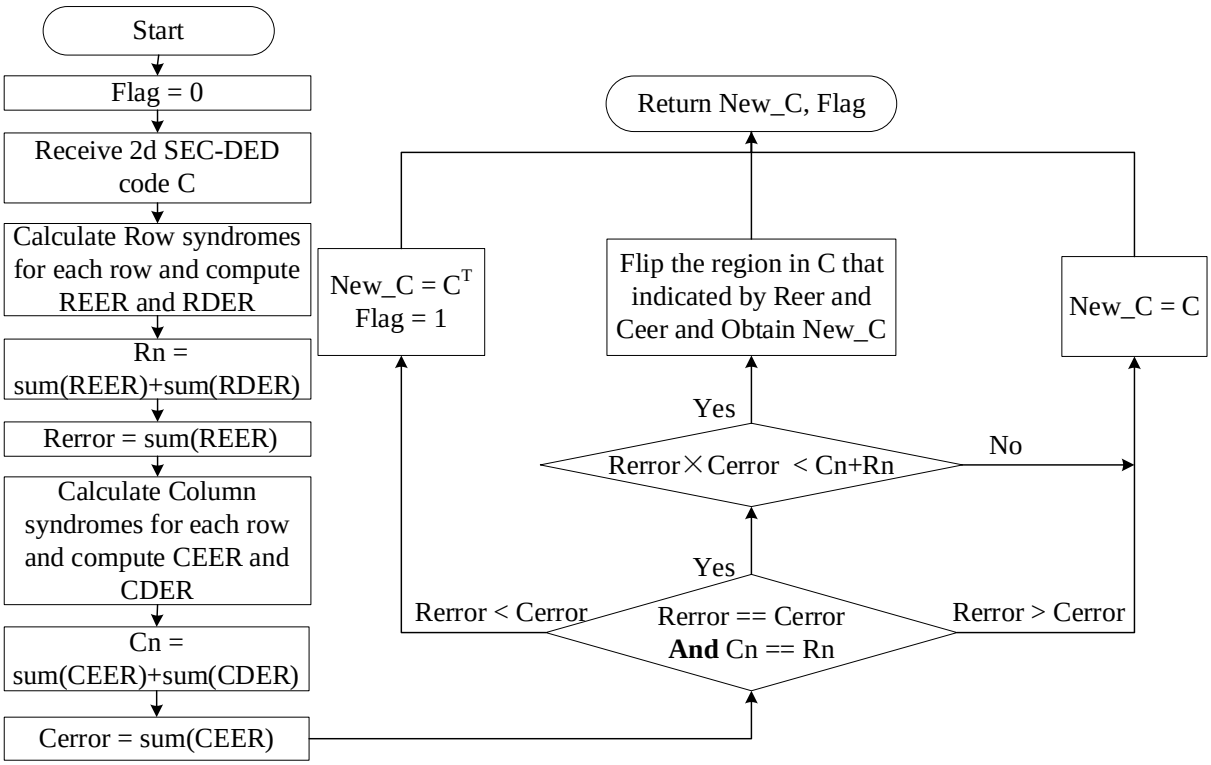


Figure 1. Flow chart of the preprocessing procedure

2. Decoding procedure.

The four registers introduced in the previous procedure are also used in this procedure. However, since the proposed decoding procedure is a modified version of Bao's three-step iterative decoding method [5], the usage of the CDER and REER is changed to that of the row status vector and column status vector used in Bao's method.

The proposed iterative decoding procedure is a three-step decoding method.

In the first step, the row syndromes for each row are calculated, and on this basis, the REER and RCDR are updated; then, row decoding is conducted. All the correctable single errors are flipped in accordance with the syndromes.

In the second step, the column syndromes for each column are calculated, and the CEER and CDER are updated. If the number of 1 value in the RDER is equal to 3 and the number of 1 value in the CEER is equal to 2, then erasure is conducted at the coordinates indicated by the RDER and CEER. Otherwise, column decoding is conducted based on the column syndromes, followed by row decoding, in which the syndromes for each row are recalculated. Then, single errors are corrected based on these updated row syndromes, and double errors are flipped based on the CDER.

In the last step, the column decoding process is repeated to correct the remaining single errors. Then, the corrected code may be transposed in accordance with the flag generated in the previous procedure. The flow chart of the decoding procedure is presented in Figure 2.

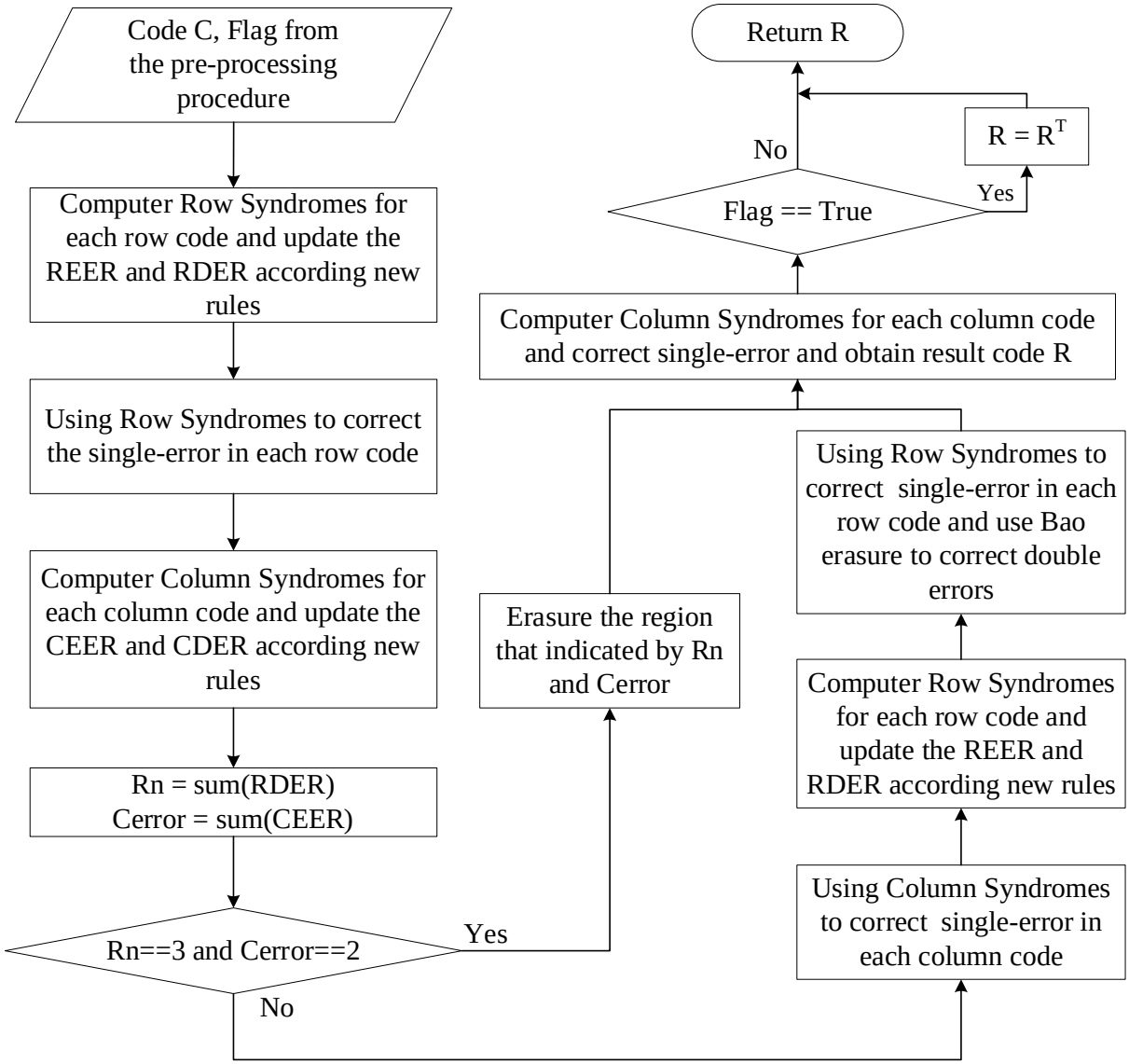


Figure 2. Flow chart of the decoding procedure

Experiment based on error patterns

To explore the potential of all implemented decoding methods, experiments based on error patterns were conducted. The number of error bits was manually set, but the error positions among the error patterns were randomly generated. It is noted that the size of the error patterns was kept the same as the size of the original (64, 16, 16) code obtained by encoding a random 16-bit binary message. The number of error bits in the error patterns was gradually increased from one error to twelve errors. For the error patterns with no more than 5 errors, we generated all possible error patterns and then added them to the codeword separately and attempted to correct these errors using each implemented decoding method. For the error patterns with more than 5 errors, since generating all the error patterns would be very difficult and time consuming, we randomly selected one million samples from all error patterns with a given number of errors for the decoding experiments. Table 1 shows the numbers of error patterns used in the current experiments for different given numbers of error bits.

Table 1. **Numbers of error patterns and error bits**

Number of Error Bits	Number of Error Patterns
1	64
2	2016
3	41664
4	635376
5	7624512
6	1000000
7~9	1000000

Table 2 and Table 3 summarize the numbers of word errors and bit errors made with the different decoding methods under various numbers of error bits. Table 4 and Table 5 were obtained by normalizing the data shown in Table 2 and Table 3, respectively. Figure 3 and Figure 4 are visualizations of Table 4 and Table 5, respectively.

Table 2. **Numbers of word decoding errors under a given number of error bits**

Given Number of Error Bits	Decoding Method			
	Two-Step Method [5]	Kreshchuk's Method [10]	Bao's Method [8]	Proposed Method
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	10192	0	0	0
5	58208	18816	0	0
6	191112	24974	383	0
7	369766	63753	5569	0
8	578553	138496	32585	3229
9	770939	254869	116425	25084

Table 3. **Numbers of bit decoding errors under a given number of error bits**

Given Number of Error Bits	Decoding Method			
	Two-Step Method [5]	Kreshchuk's Method [10]	Bao's Method [8]	Proposed Method
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	21952	0	0	0
5	1229312	75264	0	0
6	445510	159180	1655	0
7	949792	745824	23714	0
8	1714958	2171274	140275	17828
9	2765042	5035850	524295	141308

Table 4. **Normalization of the word errors produced by the decoding methods under a given number of errors bits**

Given Number of Error Bits	Decoding Method			
	Two-Step Method [5]	Kreshchuk's Method [10]	Bao's Method [8]	Proposed Method
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0,01604	0	0	0
5	0,00763	0,00246	0	0
6	0,19111	0,02497	0,00038	0
7	0,36976	0,06375	0,00556	0
8	0,57855	0,13849	0,03258	0,00322
9	0,77093	0,25486	0,11642	0,02508

Table 5. **Normalization of the bit errors produced by a decoding method under a given number of errors bits**

Given Number of Error Bits	Decoding Method			
	Two-Step Method [5]	Kreshchuk's Method [10]	Bao's Method [8]	Proposed Method
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0,00053	0	0	0
5	0,00251	0,00015	0	0
6	0,00696	0,00248	0,00002	0
7	0,01484	0,01165	0,00037	0
8	0,02679	0,03392	0,00219	0,00027
9	0,04320	0,07868	0,00819	0,00221

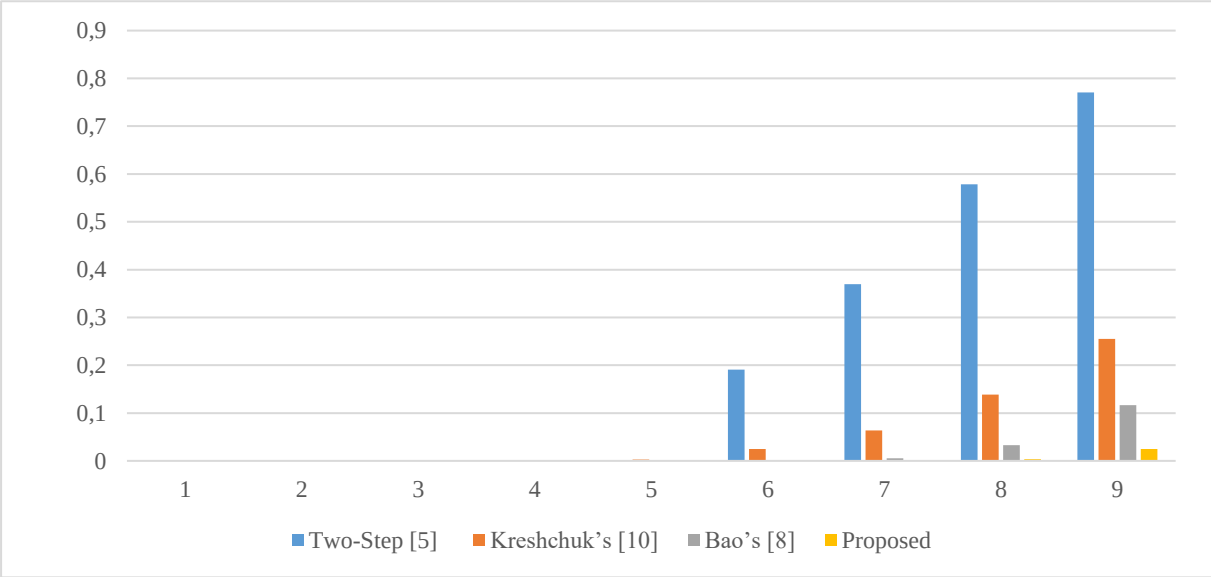


Figure 3. Histogram of the normalized word errors produced by the different decoding methods

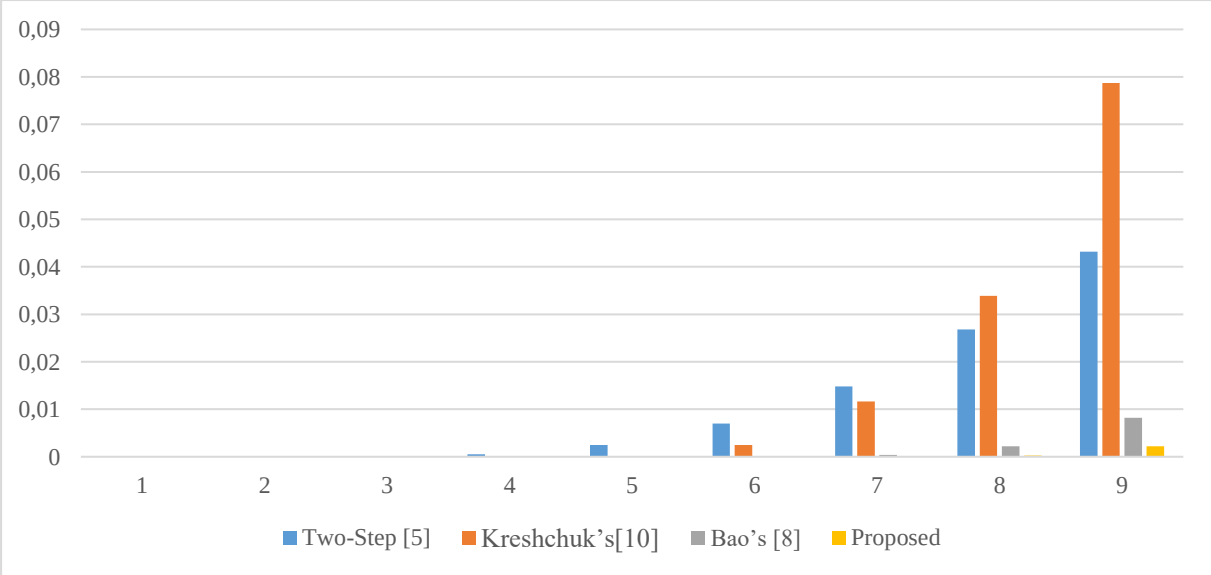


Figure 4. Histogram of the normalized bit errors produced by the different decoding methods

From the above tables, the proposed method displays the best performance in terms of the error correction capability within the range of half the minimum distance of the code, and it can properly correct more error patterns than the other three decoding methods. Exhaustive experiments have proven

that the proposed method can correct all error patterns with the number of errors below 5. For error patterns with 6 or 7 errors, the proposed method properly decodes all sampled error patterns and corrects all the error bits; therefore, there is a high probability that the proposed method also can correct all error patterns in which the number of errors is not above seven, i.e., half the minimum distance of the current code. In contrast, the two-step method, Kreshchuk's method, and Bao's method begin to produce decoding errors when the given number of errors is four, five and six, respectively.

Moreover, when the given number of errors is greater than half the minimum distance of the code, the proposed method still provides more powerful rectification ability than the other methods. For example, when the given number of errors is 9, the proposed method produces errors for only approximately 2,5 % of words and 0,02 % of bits. In contrast, the two-step method generates a 77 % word error and a 4,3 % bit error, Kreshchuk's method generates a 25 % word error and a 7,8 % bit error, and Bao's method generates an 11 % word error and a 0,8 % bit error.

Conclusion

In this paper, we have proposed an improved hard-decision iterative decoding method for 2D SEC-DED codes. The error correction capability of the proposed method is very close to half the minimum distance of the code. The decoding experiment based on a given number of errors indicated that the proposed method achieves better performance than the other decoding methods in the sense that it can correct more error patterns.

References

1. Elias P. // IEEE Trans. on Information Theory. 1954. Vol. 4. P. 29–37.
2. Forney G. // IEEE Trans. on Information Theory. 1966. Vol. 12. P. 125–131.
3. Abramson N. // IEEE Trans. on Communication Technology. 1968. Vol. 16. P. 398–402.
4. Reddy S., Robinson J. // IEEE Trans. on Information Theory. 1972. Vol. 18. P. 182–185.
5. Lin. S., Costello D.J. // Error Control Coding. Lebanon, 2001.
6. Bo F., Ampadu P. // IEEE SOC Conference. 2008. P. 59–62.
7. Bo F., Ampadu P. // VLSI Design. 2008. P. 1–14.
8. Bo F., Ampadu P. // IEEE Trans. on Circuits and Systems. 2009. Vol. 56. P. 2042–2054.
9. Kim J., Jee Y. // Proc. of International Conf. on Computer Technology and Development. 2010. P. 611–615.
10. Alexey K., Victor Z., Eygene R. // Proc.of International Workshop on Algebraic and Combinational Coding Theory. 2014. P. 211–214.
11. Blomqvist F. // Applicable Algebra in Engineering, Communication and Computing. 2021. P. 1–18
12. Chlaab A.K., Flayyih W.N., Rokhani F.Z. // Bulletin of Electrical Engineering and Informatics. 2020. Vol. 9. P.1979–1989.
13. Ren X.H., Ma J., Tsviatkou V.Yu., Kanapelka V.K. // Engineering Letters. 2022. Vol. 30. P. 948–954.

UDC 621.391

NR LDPC CODES IN 5G MOBILE COMMUNICATION SYSTEM

F.Y. LIU, S.B. SALOMATIN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 10, 2023

Abstract. In order to meet the new communication requirements and achieve low latency, high speed and high reliability connections between mobile devices, the Fifth-Generation (5G) mobile communication system. The fifth-generation (5G) mobile communication system introduces new error correction coding techniques in the data channel and control channel. Low-Density Parity-Check (LDPC) codes have been identified as the standard for 5G due to their excellent performance.has been identified as the data channel coding scheme in the 5G standard due to its excellent performance. In this paper, we introduce the construction method of LDPC code in 5G standard and simulate its decoding performance.

Keywords: 5G mobile communication, LDPC code, confidence propagation decoding.

Introduction

So far, four generations of mobile communication systems have been developed. System has a peak downlink rate of 1 Gb/s and a peak uplink rate of 500 Mb/s. The first four generations of mobile communication systems have met most of the needs of human-to-human communication. However, with the rapid development of mobile Internet, IoT and Telematics, in addition to the demand for high data rate, the demand for low latency, low power consumption and high reliability has become a new challenge for 5G mobile communication systems. The International Telecommunication Union Radiocommunications Standardization Sector has identified three major application scenarios for future 5G networks: Enhanced Mobile Broadband (eMBB), Ultra-Reliable LowLatency Communications (URLLC), and Massive Machine Communication (MMC). Machine Type Communications (mMTC).

Machine Type Communications (mMTC) [1-3]. Compared to 4G LTE (Long Term Evolution) network, the transmission rate of 5G network is 10~100 times higher; the user experience rate is 0,1~0,1. The user experience rate reaches 0,1~1 Gb/s; the latency is reduced by 5~10 times; the connected device density is increased by 10~100 times. Equipment density to improve 10 to 100 times, reaching millions per square kilometer; traffic density to improve 10 to 1000 times, to reach every square kilometer 10 ~ 1000 times, to reach tens of terabits per second per square kilometer; mobility The mobility should reach more than 500 km/h to achieve a good user experience in the high-speed railway environment.

Advantages of LDPC codes in 5G

In order to meet the needs of 5G communication, 5G New Radio (NR) adopts many new transmission technologies such as non-orthogonal multiple access, large-scale array antennas, and new channel coding techniques. Compared with 4G mobile communication system, 5G mobile communication system adopts a new pair of data channel and control channel respectively. The 5G mobile communication system adopts a pair of new channel coding techniques for the data channel and control channel, respectively. Specifically, low-density parity Low-Density Parity-Check (LDPC) code replaces the Turbo code for the data channel and the Polarization code.

Turbo codes for the data channel and polarization codes for the control channel instead of the bite-tailed convolutional codes. LDPC codes were originally proposed by Dr. Gallager, but did not receive much attention at that time due to hardware constraints. The LDPC code was originally proposed

by Dr. Gallager, but did not receive much attention at that time due to hardware limitations. It was not until the mid-1990s, with the rapid development of hardware technology, that LDPC codes were again introduced. The LDPC code was originally proposed by Dr. Gallager, but did not receive much attention at that time due to hardware limitations. Currently, the LDPC codes have been adopted by several IEEE standards, such as IEEE 802.16e, IEEE 802.11n, IEEE 802.11ac, etc. Compared with Turbo codes in 4G LTE networks, 5G NR LDPC codes have the following advantages:

1. Better area throughput efficiency and higher peak throughput.

2. Short decoding delay due to low decoding complexity and highly parallelized implementation. The advantage is more obvious at high code rates.

3. Better decoding performance for all code lengths and rates, with an error

The Frame Error Rate (FER) [4-5] is close to or below 10-5 for all code lengths and rates.

These advantages of NR LDPC codes are particularly suitable for the ultra-high throughput of 5G networks (20 Gb/s peak downlink rate and 10 Gb/s peak uplink rate) and URLLC requirements.

NR LDPC code structure in 5G

The NR LDPC coding process in 5G is shown in Figure 1. The whole NR LDPC coding chain includes code block partitioning, Cyclic Redundancy Check (CRC), LDPC coding, number-rate matching and system bit-first interleaver. First, the CRC), LDPC coding, number rate matching and system bit-first interleaver are performed. First, the large transmission block is sliced and divided into several small data blocks suitable for processing by the LDPC coder. The CRC checksum combined with the inherent error detection capability of the Parity Check Matrix (PCM) of the LDPC code can achieve a very low probability of error miss. The CRC checksum combined with the inherent error detection capability of the LDPC Parity Check Matrix (PCM) can achieve a very low probability of error misses. Again, the data block is encoded with LDPC. Then, in order to match the carrying capacity of the channel and achieve the required bit rate, a rate matching process is performed; including punching and finally, the data block is retransmitted after a system bit that enables more reliable transmission of the system bits than the checksum bits. Finally, the final encoded bits are obtained by a system bit-first interleaver that enables more reliable transmission of system bits than check bits.

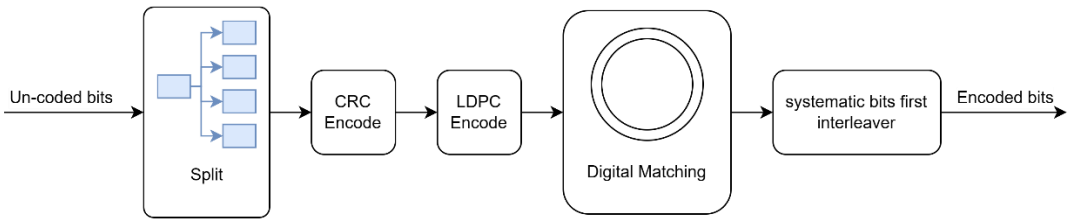


Figure 1. 5G NR LDPC encoding flow chart

NR LDPC code is a quasi-cyclic LDPC code, whose PCM is constructed by a small basic matrix. Z denotes the sub-block size and each element of the basic matrix represents a square matrix of size $Z \times Z$. Z denotes the sub-block size and each element of the basic matrix represents a square matrix of size $Z \times Z$. This square matrix can be an all-zero matrix of $Z \times Z$, or a unitary matrix of $Z \times Z$. This matrix can be a $Z \times Z$ all-zero matrix or a $Z \times Z$ unitary matrix cyclically shifted to the right by a number of bits. The number of cyclic right shifts is determined by the corresponding shift factor in the basic matrix.

Two basic matrices of NR LDPC codes

In 5G networks, the data channel can support two basic matrices, and to ensure In order to ensure good performance and low decoding delay, the standard gives the range of message block length and code rate for the two basic matrices. In order to ensure good performance and low decoding delay, the standard gives the range of message block length and code rate for the two basic matrices, and the specific parameters are shown in Table 1. Basic matrix 1 is mainly for large message blocks and high code rate, as can be seen from Table 1, the maximum message block length of basic matrix 1 can reach 8448, and the highest code rate can be 8448. Basic matrix 2 is designed for small message blocks

and low code rates. The minimum block length is only 308 and the minimum code rate is only the minimum block length is only 308 and the minimum code rate is only 1/5, which is much lower than the Turbo code in LTE. Because NR LDPC codes can use very low code rate to obtain additional coding gain, NR LDPC codes can be used in scenarios requiring high reliability. in scenarios that require high reliability.

Table 1. **Basic matrix parameters of NR LDPC codes**

Matrix parameters	Basic Matrix 1	Basic Matrix 2
Range of design code rates	$1/3 \sim 8/9$	$1/5 \sim 2/3$
Number of rows of the basic matrix	46	42
The number of columns of the basic matrix	68	52
Range of design code lengths	$308 \sim 8448$	$40 \sim 3840$
Number of non-zero elements	316	197

From Table 1, it can be seen that there is a clear overlap in the information block size and code rate of both basic matrices, which means that both basic matrices can be used in this range matrices. However, the two basic matrices have different performance for the same block size and code rate. We generally use the best fundamental matrix with different performance.

From the decoding complexity point of view, for a given block size of information, using the basic matrix2 works better for a given block size because it is more compact. Usually, the decoding delay is proportional to the number of non-zero elements in the base the number of non-zero elements in the matrix is usually proportional to the number of non-zero elements in the matrix. As can be seen from Table 1, for a given code rate, the number of non-zero elements in the basic matrix 2 is much smaller than that of the basic matrix 1, e.g. at code rate 1/3, the number of non-zero elements of basic matrix 2 is about 0,38 of that of basic matrix 1. This means that the decoding delay of basic matrix 2 has a significant decrease compared to basic matrix 1.

Table 1 shows the regional ranges of code rates and message block sizes corresponding to the two basic matrices. Typically, basic matrix 2 is used for low code rates and basic matrix 1 for high code rates. The information block size is represented by the parameter K and the code rate is represented by the parameter R. When $K \le 308$, only the basic matrix 2 can be used because in this information block size range, the basic matrix 2 has better decoding performance at all code rates compared to the basic matrix 1. When $308 \le K \le 3840$, since the code rate range of basic matrix 2 is, basic matrix 2 can reach $2/3$ in this information block range. for basic matrix 2, code rates higher than $2/3$ can be achieved by punching, but at code rates, basic matrix 1 has better decoding performance. When $K > 3840$ and, the decoding performance of basic matrix 2 is better. basic matrix 1 needs to be combined with repetitive coding to achieve the code rate, so basic matrix 2 is chosen.

5G NR LDPC code performance simulation

In order to compare the performance of NR LDPC codes composed of two basic matrices, we have performed extensive simulations for different code lengths and code rates. The decoding algorithm used in this paper is a soft-judgment decoding algorithm in a binary additive Gaussian white noise channel, Belief Propagation (BP) algorithm. The Belief Propagation algorithm updates the state information of each node by passing information from node to node, and this algorithm is an iterative approach. After several iterations, the information of all nodes no longer changes, and then the final result is obtained by judgment.

Assume that the original message symbol is $S = \{s_1, s_2, \ldots, s_k\}$, After the LDPC encoder produces n LDPC coded symbols are generated and the coded symbols are modulated with BPSK $(0 \to 1, 1 \to -1)$, Get the symbol to be sent $X = \{x_1, x_2 \ldots, x_n\}$, then after the Gaussian channel is transmitted, and the final symbol received at the receiver is:

$$y_i = x_i + n_i . \quad (1)$$

The x_i is the symbol after modulation $x_i \in \{-1,1\}$, n_i is a Gaussian random variable, $n_i \sim N(0,\sigma_n^2)$ and $\sigma_n^2 = N_0/2$, $N_0/2$ is the bilateral power spectral density of Gaussian white noise. The transmitted information is measured by the Log Likelihood Ratio (LLR). The encoding process of LDPC code is to continuously establish a linear relationship between the information symbols and the encoding symbols. This linear relationship can be expressed in terms of the encoding matrix, which is also known as the basic matrix. Suppose that after *l* iterations, $L^l_{v_i \to c_j}$ denotes the information passed from check node *j* to variable node *i*, $L^l_{v_i \to c_j}$ denotes the variable node *i* to the check *j* node. The specific steps of the decoding process are as follows: First, the LLR value from the channel is calculated and the LLR value of the channel is used as the initial value of the iteration of the variable node:

$$L(x_i \mid y_i) = \ln[\frac{P(x_i = +1 \mid y_i)}{P(x_i = -1 \mid y_i)}] = \frac{2}{\sigma^2} y_i . \tag{2}$$

The external messages are continuously exchanged between the variable node and the check node, and the message update rule from the check node to the variable node is shown in equation (3). The message update rule from the check node to the variable node is shown in equation (3), where denotes the set of all denotes the set of all variable nodes connected to the check node *j*; the message update rule from the variable node to the check node is shown in Eq. The message update rule from the check node to the variable node is shown in equation (4), where similarly denotes the set of all check nodes connected to the variable node *i*. denotes the set of all check nodes connected to variable node *i*.

$$L^l_{c_j \to v_i} = 2\tanh^{-1}\{\prod_{i' \in N(j)\setminus j} \tanh[\frac{1}{2} L^{l-1}_{v_i \to c_j}]\}, \tag{3}$$

$$L^l_{v_j \to c_j} = L(x_i \mid y_i) + \sum_{j' \in N(i)\setminus j} L^l_{c_j \to v_j} . \tag{4}$$

The soft information output of all variable nodes is calculated and adjudicated according to equation (5). If $q_i > 0$, then $v_i = 0$; otherwise $v_i = 1$. The decoding ends when the maximum number of iterations is reached or the checksum constraint is satisfied in advance.

$$q_i = L(x_i \mid y_i) + \sum_{j \in N(i)} L^l_{c_j \to v_i} . \tag{5}$$

Comparing the two decoding algorithms, we give some simulations as in Figure 2 and 3. Assuming the channel is a Gaussian white noise channel, using BPSK modulation, with a maximum number of 50 iterations and the energy is fully normalized and the signal-to-noise ratio is defined as in dB. It can be seen from the figure that the decoding performance of the BP algorithm is better than that of the MS algorithm.

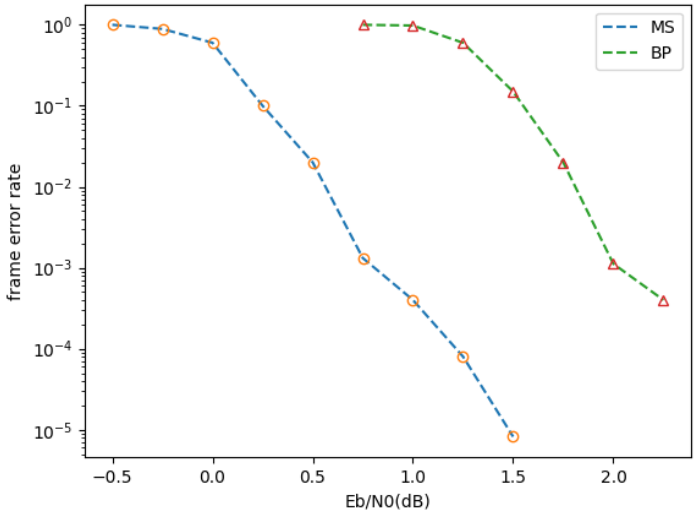


Figure 2. Simulation diagram of basic matrix 1

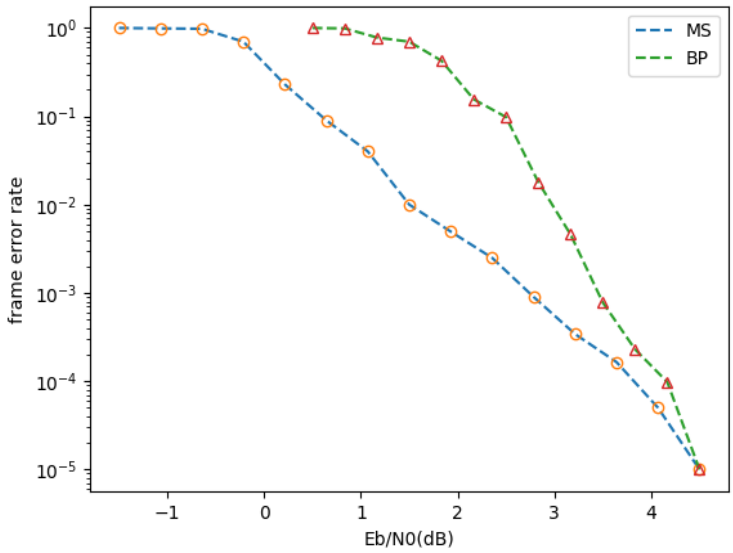


Figure 3. Simulation diagram of basic matrix 2

For example, gain for $FER = 10^3$, the dB of the BP algorithm is greater than that of the MS algorithm, which can withstand stronger noise interference, so the BP algorithm performs better than the MS algorithm.

Conclusion

In this paper, we introduce a new channel coding scheme, namely NR LDPC code in 5G. First, we give the whole flow of NR LDPC coding, and describe the uses and related operations of the key steps in the whole flow. Then, we compare the various parameters of the two basic matrices in 5G data channels. Then, we compare the various parameters of the two basic matrices in the 5G data channel; finally, we introduce in detail the decoding algorithms of the two LDPC codes, BP and MS. Finally, two decoding algorithms for LDPC codes, BP and MS, are introduced in detail.

References

1. Nam Y., Young K. // Proceedings ISCC 2000 Fifth IEEE Symposium on Computers and Communications. 2000. P. 732–737.
2. Bo F., Ampadu P. // 2008 IEEE International SOC Conference. 2008. P. 59–62.
3. Antipolis S. // Tech Rep. 2017. P. 504–507.
4. Vila Casado A., Wesel D. // IEEE Transactions on Communications. 2010. P. 3470–3479.
5. Richardson J., Shokrollahi M. // IEEE Transactions on Information Theory. 2001. P. 619–637.

УДК 621.391

ФОРМИРОВАНИЕ КОМБИНИРОВАННЫХ АСМ-ИЗОБРАЖЕНИЙ НА ОСНОВЕ ВЗВЕШЕННОГО СЛОЖЕНИЯ ДВУХ КОМПОНЕНТ

М.Ю. ЛОВЕЦКИЙ[1,2], В.Ю. ЦВЕТКОВ[2], А.А. БОРИСКЕВИЧ[2], И.И. ЛЕВОНЕНКО[2], В.А. ЛАПИЦКАЯ[1], С.А. ЧИЖИК[1]

1 – Институт тепло- и массообмена имени А.В.Лыкова НАН Беларуси, Республика Беларусь
2 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Рассматривается задача взвешенного сложения компонентных изображений атомного силового микроскопа (АСМ). Получены зависимости локальной корреляционной метрики от размера окна корреляционного анализа и вклада компонентных АСМ-изображений в результирующее комбинированное АСМ-изображений. Предложена схема адаптивного взвешенного сложения компонентных АСМ-изображений.

Ключевые слова: атомная силовая микроскопия, оценка качества комбинирования изображений, корреляция изображений, взвешенное сложение изображений.

Введение

Атомно-силовая микроскопия (АСМ) поверхности материала использует несколько параллельных синхронизированных измерительных каналов для различных физических величин (высоты, вязкости и жесткости поверхности, деформации зонда и рассеивания энергии). Формируемые в этих измерительных каналах значения компонуются в несколько двухмерных матриц чисел, представляемых многоканальными АСМ-изображениями, в которых яркости пикселей каждого канала отражают значения измеряемой физической величины в соответствующих точках поверхности. Для эффективного визуального анализа многоканальных АСМ-изображений необходимо объединять их каналы для отображения на стандартных мониторах, имеющих относительно узкий динамический диапазон, с минимальными искажениями и потерями деталей. В данной работе рассматриваются комбинированные полутоновые АСМ-изображения на основе двух измерительных каналов.

Для объединения изображений используются подходы, основанные на взвешенном сложении, методе главных компонент [1], дискретном вейвлет-преобразовании [2], однако они специально не ориентированы на объединение изображений, формируемых в измерительных каналах атомного силового микроскопа. Отсутствуют рекомендации по выбору алгоритма для эффективного формирования комбинированных АСМ-изображений. Для их разработки необходима оценка качества комбинированных изображений. Известные показатели качества изображений основаны на анализе краев [3, 4], взаимной информации [5], оценке количества информации в изображении [6], оценке точности визуальной информации в различных масштабах представления изображения [7], однако они специально не ориентированы на оценку качества комбинированных АСМ-изображений. Относительной простотой вычислений отличается коэффициент корреляции, использующий средние значения изображений, но не учитывающий локальные особенности распределения яркости. Для оценки качества АСМ-изображений, отличающихся существенными локальными неоднородностями распределения яркости, представляет интерес метрика качества комбинирования компонентных АСМ-изображений на основе коэффициентов локальной корреляции, учитывающая вклад каждого из компонентных АСМ-изображений в результирующее комбинированное АСМ-изображение и корреляцию между компонентными АСМ-изображениями. Локальная

корреляция обеспечивает более высокую точность оценки качества комбинирования АСМ-изображений по сравнению с глобальной корреляцией, но ее значения зависят от размера окна анализа.

Целью работы является определение вкладов значений пикселей компонентных АСМ-изображений, обеспечивающих передачу в комбинированное АСМ-изображение наиболее полной информации об объектах определенного размера.

Формирование комбинированных АСМ-изображений

Исходя из предположения о независимости эффективности методов объединения изображений и точности показателей качества комбинированных изображений для формирования комбинированных АСМ-изображений выбран простейший метод взвешенного сложения (рис. 1). Согласно данному методу значения пикселей $m_C(y,x)$ комбинированного АСМ-изображения $M_C = \left\| m_C(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ вычисляются на основе значений пикселей АСМ-изображений $M_1 = \left\| m_1(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ первого и $M_2 = \left\| m_2(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ второго измерительных каналов атомного силового микроскопа с помощью выражения

$$m_C(y,x) = \left[k\, m_1(y,x) + (1-k)\, m_1(y,x) \right] \tag{1}$$

при $y = \overline{0,Y-1},\ x = \overline{0,X-1}$,

где k – коэффициент, определяющий вклад значений пикселей каждого компонентного АСМ-изображения M_1 и M_2 в значения пикселей комбинированного АСМ-изображения M_C, $0 < k < 1$; Y, X – размеры (в пикселях) компонентных и комбинированного АСМ-изображений по вертикали и горизонтали; [ ] – операция округления значений пикселей до ближайшего целого.

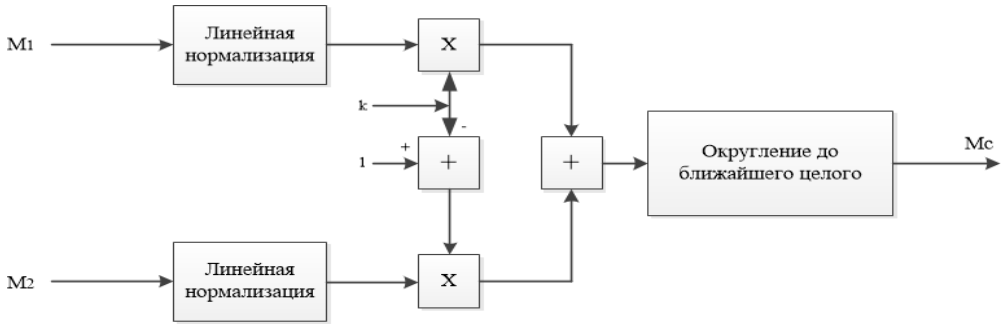


Рис. 1. Схема формирования комбинированного АСМ-изображения на основе взвешенного сложения компонентных АСМ-изображений

Меньшие значения коэффициента k на рис. 1 соответствуют меньшей относительной доли значений компонентного АСМ-изображения M_1 в комбинированном АСМ-изображении M_C по сравнению с компонентным АСМ-изображением M_2.

Оценка качества комбинирования АСМ-изображений на основе коэффициента локальной корреляции

Повышение точности корреляционной оценки качества комбинирования АСМ-изображений достигается за счет учета локальных особенностей распределений значений пикселей в компонентных и комбинированном АСМ-изображениях. Для этого используется коэффициент $r_{\mathrm{L}}(A,B)$ локальной корреляции двух АСМ-изображений $A = \left\| a(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ и $B = \left\| b(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$, вычисляемый с помощью выражения

$$r_{\mathrm{L}}(A,B,p)=\frac{\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left|a(y,x)-a_{\mathrm{L}}(y,x,p)\right|\left|b(y,x)-b_{\mathrm{L}}(y,x,p)\right|}{\sqrt{\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left(a(y,x)-a_{\mathrm{L}}(y,x,p)\right)^2\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left(b(y,x)-b_{\mathrm{L}}(y,x,p)\right)^2}}, \quad (2)$$

где $a_{\mathrm{L}}(y,x,p)$, $b_{\mathrm{L}}(y,x,p)$ – средние значения яркостей пикселей изображений A и B в окрестности пикселя с координатами (y,x) размером $p\times p$ пикселей, $a_{\mathrm{L}}(y,x,p)=\frac{1}{p^2}\sum_{j=0}^{p}\sum_{i=0}^{p}a(y+j,x+i)$, $b_{\mathrm{L}}(y,x,p)=\frac{1}{p^2}\sum_{j=0}^{p}\sum_{i=0}^{p}b(y+j,x+i)$.

Для оценки качества комбинирования АСМ-изображений с учетом корреляции между комбинированным АСМ-изображением и каждым из двух компонентных АСМ-изображений, а также между компонентными АСМ-изображениями в [8] предложена локальная корреляционная метрика $D_{\mathrm{L}}(k)$, вычисляемая с помощью выражения (чем больше ее значение, тем лучше)

$$D_{\mathrm{L}}(k)=\frac{r_{\mathrm{L}}(M_C,M_1,k)+r_{\mathrm{L}}(M_C,M_2,k)}{\left|r_{\mathrm{L}}(M_C,M_1,k)-r_{\mathrm{L}}(M_C,M_2,k)\right|r_{\mathrm{L}}(M_1,M_2,0,5)}. \quad (3)$$

Локальная корреляционная метрика $D_{\mathrm{L}}(k)$ позволяет определить значение k, обеспечивающее лучшее соотношение вкладов компонентных АСМ-изображений в комбинированное АСМ-изображение по сравнению с глобальной корреляционной метрикой.

Зависимости метрики качества комбинирования компонентных АСМ-изображений от размера окна корреляционного анализа

На рис. 2 приведены зависимости значений метрики $D_{\mathrm{L}}(k)$ от размера p окна корреляционного анализа и коэффициента k для 10 комбинированных АСМ-изображений. Из рис. 2 следует, что для некоторых АСМ-изображений локальные максимальные значения метрики $D_{\mathrm{L}}(k)$ зависят от значения p. Размер p окна корреляционного анализа определяет размер значимых объектов на компонентных АСМ-изображениях, которые должны вносить основной вклад в комбинированное АСМ-изображение M_C.

Из рис. 2 следует, что для АСМ-изображений 5 – 8 при любых p наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ обеспечиваются при k = 0,7. На рис. 3 приведены компонентные АСМ-изображения 5 – 8, полученные при различных значениях k. Для АСМ-изображений 1 – 4, 9, 10 наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ при различных значениях p достигаются для различных значений k.

При необходимости передачи в комбинированные АСМ-изображения 1, 3, 9, 10 наиболее полной информации о мелких объектах (p = 3) компонентных АСМ-изображений необходимо использовать значения k, равные 0,7, 0,5, 0,7, 0,5, соответственно. При необходимости передачи в комбинированные АСМ-изображения 1, 3, 9, 10 наиболее полной информации о более крупных объектах (p > 5) необходимо использовать значения k, равные 0,5, 0,3, 0,5 (0,3 при p > 40), 0,3, соответственно. На рис. 4 приведены компонентные АСМ-изображения 1, 3, 9, 10, полученные при различных значениях k.

Для АСМ-изображения 2 при p < 40 наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ обеспечиваются при k = 0,7. При p > 40 наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ обеспечиваются при k = 0,5. Для АСМ-изображения 4 при p < 110 наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ обеспечиваются при k = 0,5. При p > 110 наибольшие значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ обеспечиваются при k = 0,3. На рис. 5 приведены компонентные АСМ-изображения 2 и 4, полученные при различных значениях k.

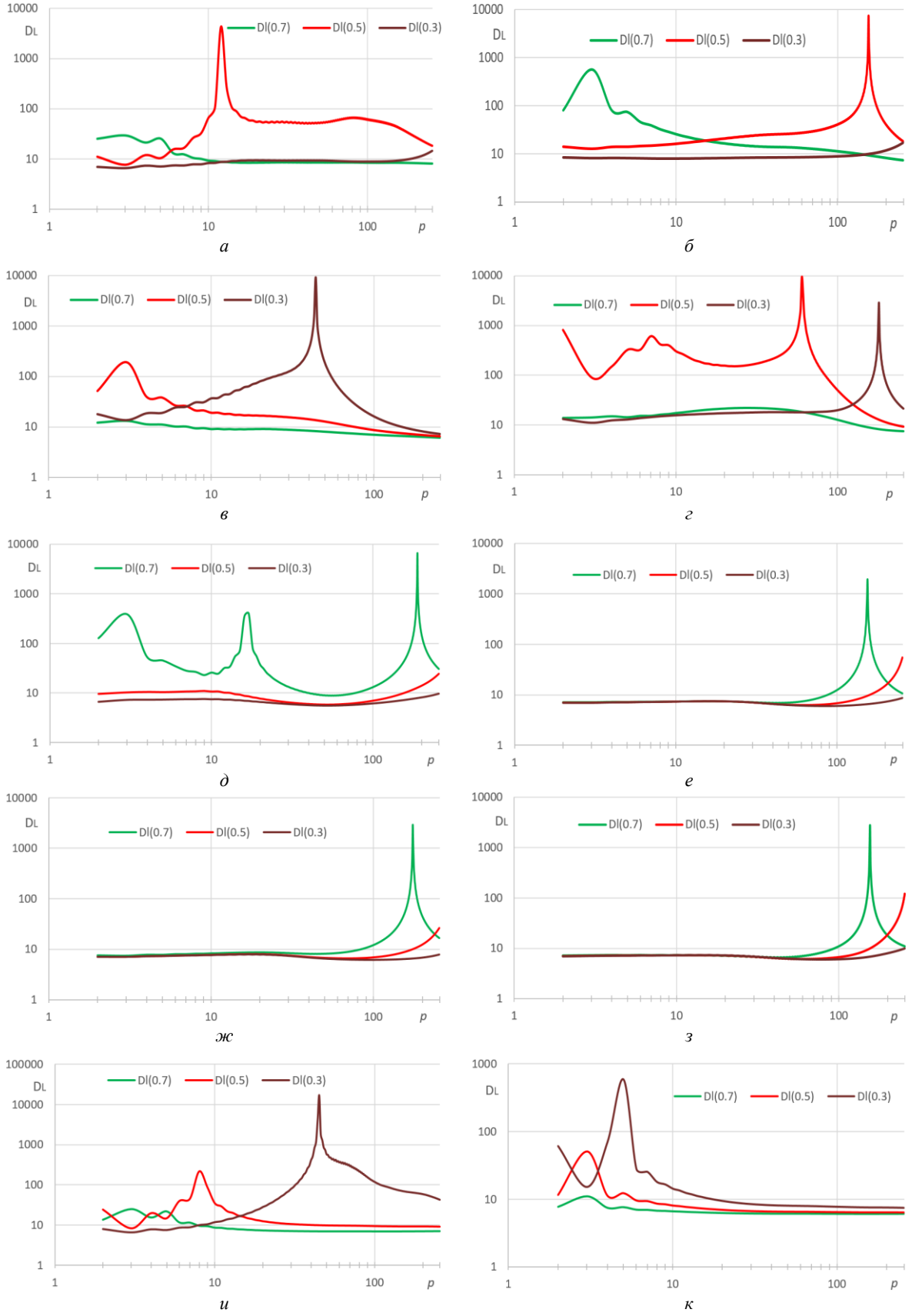
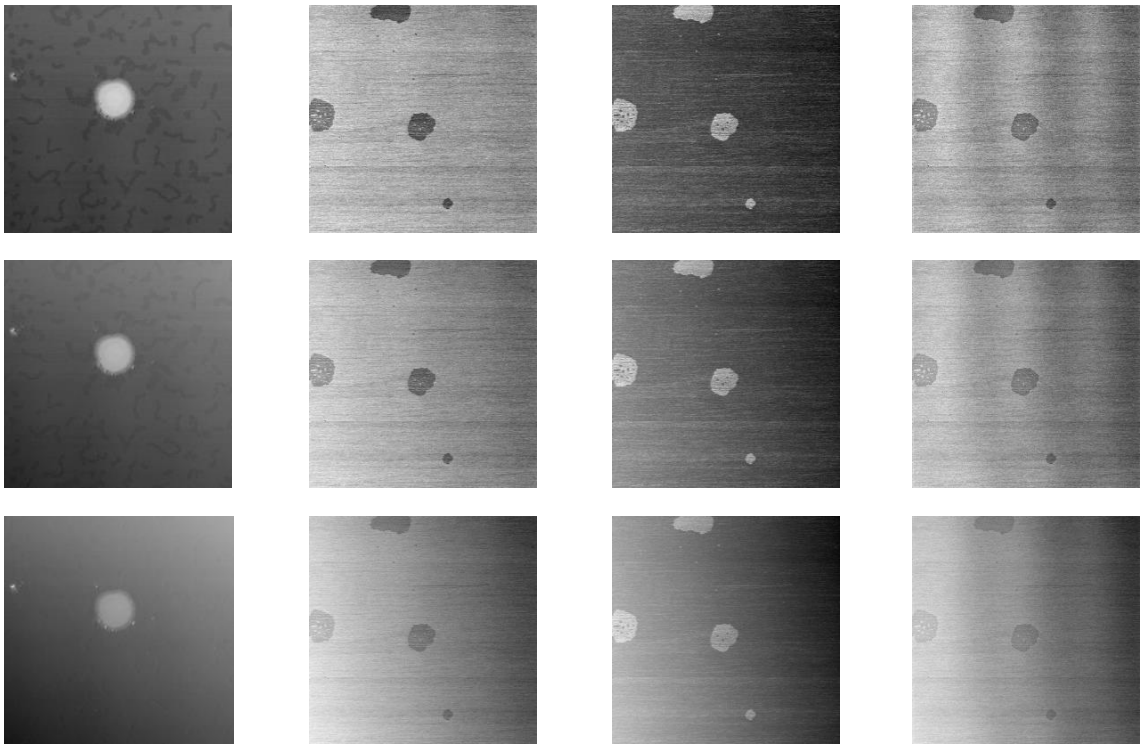
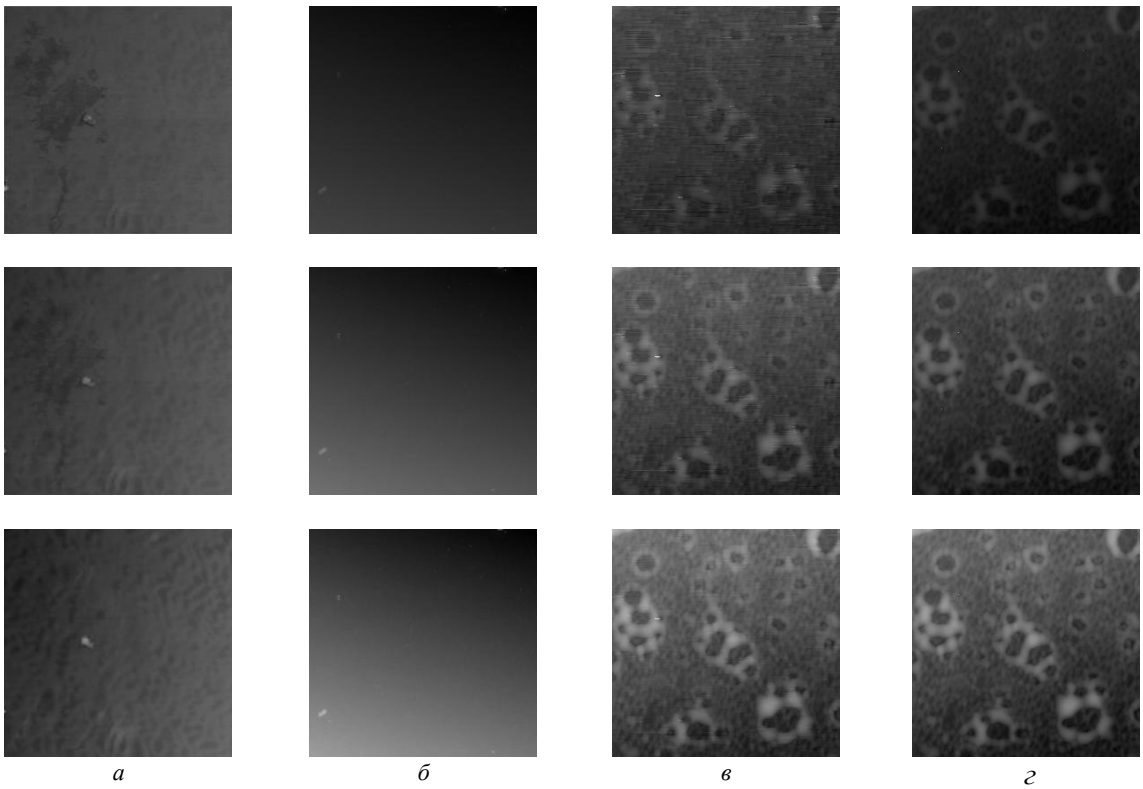


Рис. 2. Зависимости значений локальной корреляционной метрики от размера окна корреляционного анализа для компонентных АСМ-изображений: *а* – АСМ-1; *б* – АСМ-2; *в* – АСМ-3; *г* – АСМ-4; *д* – АСМ-5; *е* – АСМ-6; *ж* – АСМ-7; *з* – АСМ-8; *и* – АСМ-9; *к* – АСМ-10



а *б* *в* *г*

Рис. 3. Комбинированные АСМ-изображения при значениях k=0,7 (верхний ряд), k=0,5 (средний ряд), k=0,3 (нижний ряд): а – АСМ-5; б – АСМ-6; в – АСМ-7; г – АСМ-8

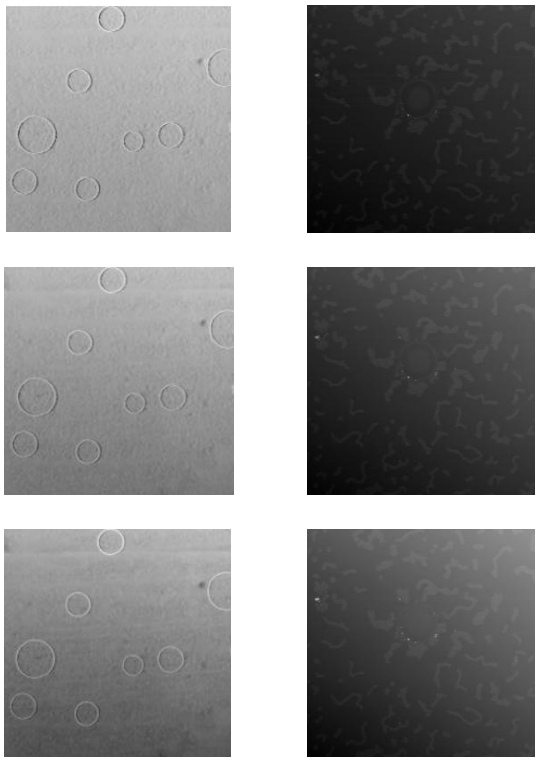


а *б* *в* *г*

Рис. 4. Комбинированные АСМ-изображения при значениях k=0,7 (верхний ряд), k=0,5 (средний ряд), k=0,3 (нижний ряд): *а* – АСМ-1; *б* – АСМ-3; *в* – АСМ-9; *г* – АСМ-10

Из рис. 2 следует, что по глобальному максимальному значению локальной корреляционной метрики $D_{\mathrm{L}}(k)$ во всем диапазоне изменения значения p можно определить значение k, обеспечивающее лучшие условия для передачи в комбинированное АСМ-

изображение информации об объектах компонентных АСМ-изображений, имеющих наиболее часто встречающиеся размеры. С учетом данного свойства предлагается схема адаптивного взвешенного сложения компонентных АСМ-изображений (рис. 6) с автоматическим выбором значения *k*, определяющим вклад значений пикселей компонентных АСМ-изображений в комбинированное АСМ-изображение.



а *б*

Рис. 5. Комбинированные АСМ-изображения при значениях k=0,7 (верхний ряд), k=0,5 (средний ряд), k=0,3 (нижний ряд): *а* – АСМ-2; *б* – АСМ-4

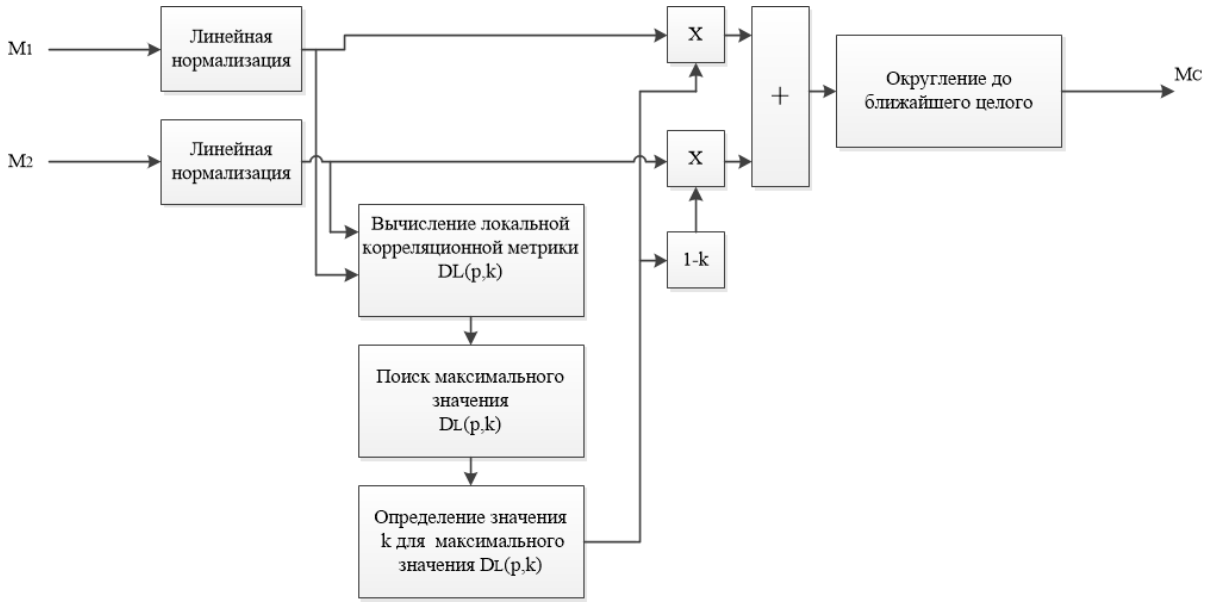


Рис. 6. Схема адаптивного формирования комбинированного АСМ-изображения на основе взвешенного сложения компонентных АСМ-изображений с автоматическим определение вклада значений пикселей компонентных АСМ-изображений в комбинированное АСМ-изображение

Схема на рис. 6 получена из схемы, приведенной на рис. 1, за счет введения дополнительных блоков, обеспечивающих вычисление значений локальной корреляционной

метрики $D_{\mathrm{L}}(k)$ для различных значений p и k с последующим поиском максимального значения $D_{\mathrm{L}}(k)$ и определением соответствующего ему значения k. Адаптивное формирование комбинированного АСМ-изображения имеет существенно более высокую вычислительную сложность по сравнению с обычным формированием. Значения локальной корреляционной метрики $D_{\mathrm{L}}(k)$ записываются в матрицу размером $p \times k$, что приводит к соответствующему росту пространственной сложности. Временная сложность увеличивается на $2pk$ операций, необходимых для вычисления значений локальной корреляционной метрики и поиска ее максимального значения.

Заключение

Получены зависимости значений локальной корреляционной метрики от размера окна корреляционного анализа и вклада компонентных АСМ-изображений в комбинированное АСМ-изображение. По данным зависимостям установлены вклады значений пикселей компонентных АСМ-изображений, обеспечивающих передачу в комбинированное АСМ-изображение наиболее полной информации об объектах определенного размера. Предложена схема адаптивного взвешенного сложения компонентных АСМ-изображений с автоматическим определением вклада значений пикселей компонентных АСМ-изображений в комбинированное АСМ-изображение.

FORMATION OF COMBINED AFM IMAGES BASED ON WEIGHTED ADDITION OF TWO COMPONENTS

M.Yu. LAVETSKI, V.Yu. TSVIATKOU, A.A. BORISKEVICH, I.I. LIAVONENKA, V.A. LAPITSKAYA, S.A. CHIZHIK

Abstract. The problem of weighted addition of component images of an atomic force microscope (AFM) is considered. The dependences of the local correlation metric on the size of the correlation analysis window and the contribution of component AFM images to the resulting combined AFM images were obtained. A scheme for adaptive weighted summation of component AFM images was proposed.

Keywords: atomic force microscopy, image combination quality assessment, image correlation, weighted image summation.

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. Petrovic V., Xydeas C. // Tenth IEEE International Conference on Computer Vision (ICCV'05). 2005. Vol. 1, P. 1866-1871.
4. Piella G., Heijmans H. // Proceedings International Conference on Image Processing (Cat. No.03CH37429). 2003. P. 111-173.
5. Qu G., Zhang D., Yan P. // Opt. Express. 2001. Vol. 9. P. 184-190.
6. Aslantas V., Bendes E. // AEU – International Journal of Electronics and Communications. 2015. P. 1890-1896.
7. Han Y., Cai Y., Cao Y., Xu X. // Inf. Fusion. 2013. Vol. 14. No. 2. P. 127–135.

UDC 611.1; 519.25

HUMAN HEART RATE MONITORING BASED ON FACIAL VIDEO PROCESSING

N.V. BACH, I.A. BORISKIEVIC

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 23, 2023

Abstract. Heart rate (HR) is one of the most important physiological parameters and a vital indicator of people's physiological state, making it important to monitor. Over the last decade, research has focused on non-contact systems, which are simple, low-cost, and comfortable to use. This paper analyses the complexity of each step in the development of a human heart rate monitoring algorithm based on facial video processing. Specifically, the research focuses on the pulse signal extraction step. The proposed algorithm based on the transform of 2D signal to 1D signal, its detrending and window discrete transform are used to improve the accuracy of HR estimation. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2 bpm.

Keywords: heart rate, facial video processing, remote photoplethysmography, window discrete transform.

Introduction

The human pulse is a rhythmic oscillation of the vessels that correspond to the contractions of the heart and it is one of the most important indicators that helps to track whether everything is good with the heart. Traditional heart rate detection mainly includes two ways: electrocardiograph (ECG) and contact photoplethysmography (cPPG) based on sensors. Due to the limitations of cPPG methods, it is particularly important to study a non-contact HR detection method. The rPPG (remote photoplethysmography) has been proven to be superior because it is non-intrusive. It may be suitable for continuous measurement of heart rate (HR) in many cases, such as neonatal ICU (intensive care unit) monitoring [1], burn victims, driver status assessment [2], online learning [3], provide low-cost solutions for health monitoring applications, another application of rPPG in health monitoring include blood perfusion mapping [4] and monitoring regional anesthesia effectiveness [5]. Although this kind of methods may not be as accurate as an electrocardiogram (ECG) device, they can provide a long-term HR monitoring without being uncomfortable for patients. These technologies can be very helpful in increasing fields like telemedicine, where usability is a key factor.

The choice of a facial ROI (region of interest) acts as the first key step of the system. First, the pulsatile signal strength varies at different locations on the face due to the distribution of capillaries beneath the skin surface. The location of an ROI has a direct impact on the quality of the raw rPPG measurement. Second, the shape of an ROI always leads to unnecessary inclusion of undesired pixels like eyes, mouth, hair, or background pixels, thus, introducing rigid/non-rigid motion artifacts. It is crucial to choose a good ROI to guarantee a higher measurement accuracy. While most rPPG approaches extract pulsed signals by averaging over all skin pixels on ROIs, we propose an algorithm that allows extracting pulse signals from ROIs for increasing HR estimation accuracy.

Our contributions can be summarized as follows:

– we present a human heart rate monitoring algorithm based on facial video processing;

– we propose an algorithm to remove outliers from signal based on Z-score method.

Human heart rate monitoring based on facial video processing algorithm

Proposed method uses video as an input and returns pulse rate as an output. Sequence of steps of the proposed algorithm can be represented as follows:

– detection of person's face on color image;
– extraction region of interest (ROI);
– transform 2D ROI to 1D ROI signal;
– processing of 1D ROI signal;
– computation of power spectrum of 1D ROI signal;
– band pass filtering;
– estimation of person heart rate.

Block diagram of the proposed algorithm to determine HR from a facial video can be represented by the graphical form.

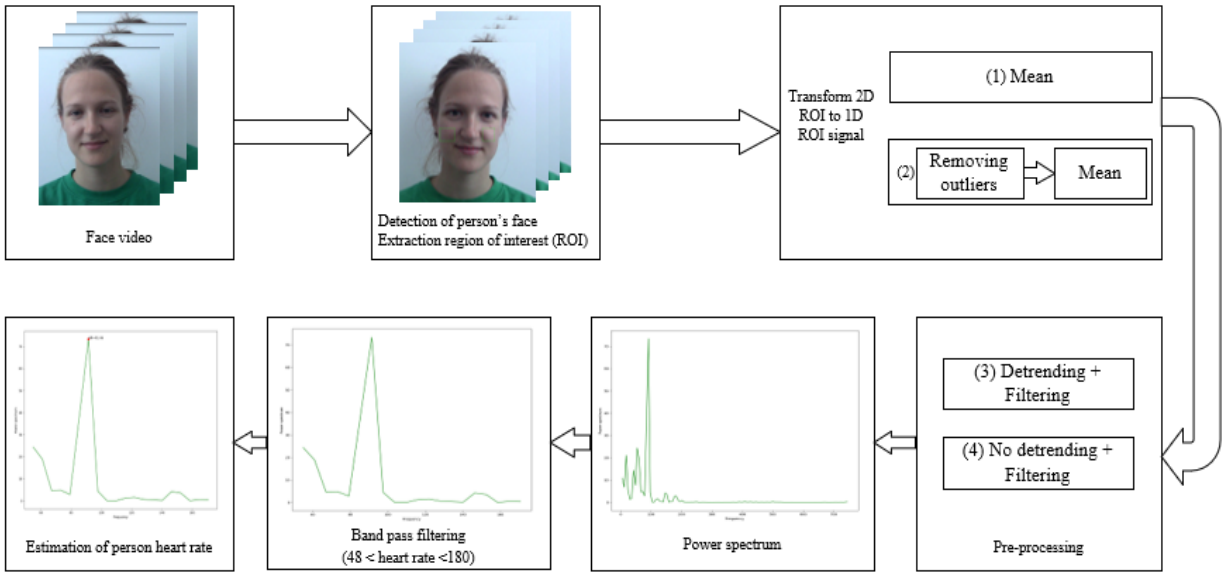


Figure 1.The sequence of steps of the proposed algorithm for estimating heart rate using facial video

Face detection is a crucial preprocessing step for the traditional rPPG methods to measure HR. Its accuracy has a direct impact on the accuracy of HR detection. At present, three mainstream methods Viola–Jones (VJ), histogram of oriented gradients (HOG) and multi-task cascaded convolutional networks (MTCNN) are often used for face detection. Dlib (an open-source library) face detector model is based on HOG feature descriptor and linear SVM (support vector machine)) classification. Dlib's HOG + linear SVM face detector is fast and efficient. Dlib HOG is the fastest method on the CPU (central processing unit) [6]. In this application we use the HOG method to locate a person face.

Extraction region of interest. The ROI in the cheeks was determined through the location of the human mouth and nose. The human eyes, mouth and nose were detected with Dlib HOG method.

Transform 2D ROIs to a 1D signal

For each frame of the video sequence, we obtain ROIs of facial skin pixels. The amplitude of the PPG-signal in light reflected from the skin varies as a function of the wavelength, showing a strong peak around 550 nm and a dip around 650 nm. Because that we use only wavelength 550 nm (green channel) to estimate HR [7].

To transform 2D ROIs to a 1D signal most rPPG approaches extract pulsed signals by averaging over all skin pixels value on ROIs. We propose algorithm to remove outliers from video frame ROI histogram based on Z-score method before extract pulsed signals.

Removal of outliers from ROI histogram based on Z-score method

The Z-score is one of the most commonly used tools in determining outliers. Z-score (Z_{score}) is just the number of standard deviations away from the mean that a certain data point is.

$$Z_{score} = \frac{I_{ROI} - \mu}{\sigma}, \quad (1)$$

where I_{ROI} – matrix of ROI pixels values, $\mu = \sum_{x=0,y=0}^{x<M,y<N} \frac{I_{ROI}(x,y)}{a \cdot b}$ and $\sigma = \sqrt{\frac{\sum_{x=0,y=0}^{x<M,y<N} (I_{ROI}(x,y) - \mu)^2}{a \cdot b}}$ – the mean value and the standard deviation of the ROI image values respectively, (*a, b)*– the width and height of video frame ROI.

To improve HR estimation accuracy expression (2) is used to detect and remove outliers from matrix of ROI pixels I_{ROI}:

$$\begin{cases} I_{ROI\,OUT} > \mu + Z_{scre} \cdot \sigma \;\; if \;\; Z_{score} < 2 \\ I_{ROI\,OUT} < \mu - Z_{scre} \cdot \sigma \;\; if \;\; Z_{score} > -2 \end{cases} \quad (2)$$

The 1D ROI signal Detrending

Detrending is an important signal processing concept, which is used to remove unwanted trend from the 1D ROI signal that represents as a sequence *s(n)* of discrete mean values of video frame ROIs. We eliminate the signal deviation trend using the adaptive iteratively re-weighted penalized least squares (Airpls) [8].

The window discrete Fourier transform

Before applying discrete Fourier transform (DFT), the PPG-signal that represents as a sequence of discrete values of mean value of video frame ROI is filtered by Hamming window $w(n) = 0,54 - 0,46\cos(\frac{2\pi n}{N-1})$, $0 \le n \le N-1$, N – the window length, n – index of discrete time samples.

The time signal of the pulse wave window is transformed into a sequence of discrete frequency samples by DFT, which is defined by:

$$S_k = \sum_{n=0}^{N-1} s_{FIL}(n) \cdot e^{-i\frac{2\pi kn}{N}}, \quad (3)$$

where S_k – k-th the DFT coefficients, $s_{FIL}(n)$ – n-th filtered value of 1D ROI signal *s(n)*, N – the number of the DFT coefficients, k ($k = 0,1,\ldots,N-1$) – frequency index.

The DFT power spectrum of 1D signal is defined as

$$P_k = \begin{cases} \frac{1}{N^2}\left|S_0\right|^2, & k = 0 \\ \frac{2}{N^2}\left|S_k\right|^2, & k = 1,\ldots,N/2, \end{cases} \quad (4)$$

An DFT was applied to the filtered pulse signal, and the heart rate was taken as the frequency where the spectral power was maximal. Then we apply band pass filter with F_l = 0.8 Hz and F_h = 3 Hz, which are 48 and 180 bpm respectively to remove unwanted frequency.

Human heart rate estimation

Human heart rate is calculated per window. The window length is set to 10s and the time distance between the two consecutive frames is equal to 0.04s.

There are usually several peaks in a same frequency domain of power spectrum. Heart rate value in the PPG signal is the position of frequency sample with the highest energy.

Experimental result

We use dataset from https://github.com/vladostan/Dataset-for-video-based-pulse-detection. Open dataset for video-based pulse detection. Includes 30 .mp4 video files and ground truth ECG signals.

Mean absolute error (MAE) and root mean square error (RMSE) [9] is selected to evaluate HR estimation accuracy by the proposed algorithm. When calculating the average error, outliers are discarded. The HR estimation accuracy based on the algorithms depicted in Figure 1, are presented in the table below.

Table 1. **Average heart rate prediction: comparison among different algorithm on different conditions with detrending (3)**

Algorithm	Normal condition			Physical activity		
	ME (bpm)	MAE (bpm)	RMSE (bpm)	ME (bpm)	MAE (bpm)	RMSE (bpm)
(1) Mean	0.21	1.83	2.16	0.75	2.04	2.61
(2) Removing outlier + mean	0.27	1.74	2.01	0.56	1.76	2.21

Table 2. **Average heart rate prediction: comparison among different algorithm on different conditions without detrending (4)**

Algorithm	Normal condition			Physical activity		
	ME (bpm)	MAE (bpm)	RMSE (bpm)	ME (bpm)	MAE (bpm)	RMSE (bpm)
(1) Mean	0.1	1.82	2.2	0.64	2.01	2.55
(2) Removing outlier + mean	0.14	1.75	2.11	0.43	1.87	2.45

It can be concluded from Tables 1 and 2 that the variant of algorithm (2), which removes outliers and applies averaging of 1D signal under different conditions with detrending, provides lower MAE and RMSE value upon estimating HR. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2bpm due to the proposed algorithms uses outlier removing procedure.

Conclusion

The proposed algorithm is based on transform 2D signal to 1D signal, removing outliers, detrending and window discrete Fourier transform. It allows us to increase the human heart rate estimation accuracy due to removing outliers from video frame ROI based on Z-score method is used. The experimental results show that the accuracy of human heart rate estimation in terms of MAE and RMSE is equal around 2bpm.

References

1. Aarts L.A.M., Jeanne V., Cleary J.P., Lieber C., Nelson J.S. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit – a pilot study // Early Hum Dev. 2013 Dec. Vol. 89, No. 12, P. 943-8.
2. Zheng K., Ci K., Cui J., Kong J., Zhou J. Non-contact heart rate detection when face information is missing during online learning // Sensors. 2020. Vol. 20, No. 24, P. 7021.
3. Taylor W., Abbasi Q.H., Dashtipour K., Ansari S., Shah S.A. A review of the state of the art in non-contact sensing for covid-19 // Sensors. 2020. Vol. 20, No. 19, P. 5665.
4. Rubins U., Erts R., and Nikiforovs V. The blood perfusion mapping in the human skin by photoplethysmography imaging // XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010. IFMBE Proceedings. 2010. Vol. 29, P. 304-306.
5. Rubins U., Miscuks A., Rubenis O., Erts R., and Grabovskis A. The analysis of blood flow changes under local anesthetic input using non-contact technique // 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China. 2010. P. 601-604.
6. Gupta V. Face Detection – Dlib, OpenCV, and Deep Learning (C++ / Python). October 2018. [Online]. Available: https://learnopencv.com/face-detection-opencv-dlib-and-deep-learning-c-python/. [Accessed February 2023].
7. De H. G., Jeanne V. Robust Pulse Rate From Chrominance-Based Rppg // IEEE Transactions on Biomedical Engineering. 2013. Vol. 60, No. 10, P. 2878 – 2886.
8. Zhang Z., Chen S., Liang Y. Baseline correction using adaptive iteratively reweighted penalized least squares // Analyst. 2010 May. Vol. 135, No. 5, P. 1138-46.
9. Pagano T.P., Santos L.L., Santos V.R. Remote Heart Rate Prediction in Virtual Reality Head-Mounted Displays Using Machine Learning Techniques // Sensors. 2022. Vol. 22, P. 9486.

UDC 004.75

VIDEO OBJECT DETECTION PROGRAM DESIGN UNDER TWO DIFFERENT CLIENT-SERVER ORGANIZATIONS

H. GAO, O.G. SHEVCHUK

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 17, 2023

Abstract. The article presents video object detection program design under two different client-server organizations based on YOLOX model. It introduces the N-tier architecture and shows how to design programs through sequence diagrams based on client-server organization cases. And the processing process of the target detection part of the program is given. The test results show that the program design can be applied to Android system and Web scene, and analyze the actual application conditions.

Keywords: software design, object detection, image preprocessing, model inference.

Introduction

Object detection is a classic task in the field of computer vision, and it is the basic premise for advanced vision tasks such as scene content analysis and understanding. The target detection task in the video is closer to the needs of real life. In real life, intelligent video surveillance, robot navigation and other application scenarios need to process the video and detect the target in the video. The target detection in the video needs to deal with the various changes of the target due to the movement on the basis of the static image target detection, which is the difficulty. With the development of deep learning, deep convolutional neural networks are rapidly applied to every field of computer vision, and have made relatively great progress compared with traditional methods. In the context of deep convolutional networks, YOLO is an advanced single-stage target detection algorithm. It has undergone the evolution of version 1 ~ version 4, and has developed to YOLOX that does not rely on anchor boxes. The proposal of YOLO aims to improve the detection efficiency of target detection, trying to make target detection reach the level of real-time detection. This article proposes two implementations of video object detection based on the YOLOX model under different client-server organizations, uses UML diagrams to describe the functions of the modules, and explains how to infer with the model. In order to verify the effectiveness, the program is implemented according to the design and tested in a real network environment.

N-tier architecture understanding

In the client-server model, a client is a piece of computer hardware or software that accesses services provided by a server as part of a computer network client-server model. The server is usually (but not always) on another computer system, in which case the client accesses the service over the network. A client can be any device – computer, tablet or mobile phone. Thus, client-server represents the relationship between collaborating programs in an application, consisting of a client that initiates a request for a service and a server that provides that function or service. There are three main categories of client-server:

1. One-tier architecture. It is the simplest one as it is equivalent to running the application on the personal computer. All of the required components for an application to run are on a single application or server.

2. Two-tier architecture. It consists of a client, a server, and a protocol that connects the two layers. The GUI code resides on the client host, and the domain logic resides on the server host. Domain

logic or business logic is the part of a program that encodes real-world business rules that determine how data is created, stored, and changed. Business logic should be distinguished from business rules. Business logic is the part of an enterprise system that determines how data is transformed or calculated, and how it is routed to people or software. Business rules are the formal expression of business policy. Anything that is a process or a procedure is business logic, and anything that is neither a process nor a procedure is a business rule. For example, welcoming a new visitor is a process consisting of steps to be taken, while saying that every new visitor must be welcomed is a business rule. In addition, business logic is procedural while business rules are declarative.

3. Multitier architecture (often referred to as an N-tier architecture) is a client-server architecture in which presentation, application processing, and data management functions are physically separated. The most widely used multi-tier architecture is the three-tier architecture. Three-tier architecture is a client-server software architectural pattern in which the user interface (presentation layer), application (functional process logic), computer data storage and data access (data layer) are developed and maintained as independent modules, usually on a different platform.

For multi-tier architecture, it offers the possibility to physically distribute client-server applications across multiple machines. For client-server organization, there is the simplest organization: a client machine containing only the programs implementing the user-interface level; a server machine containing the rest, that is the programs implementing the processing and data level. In addition, there are alternative client-server organizations, shown in Figure 1.

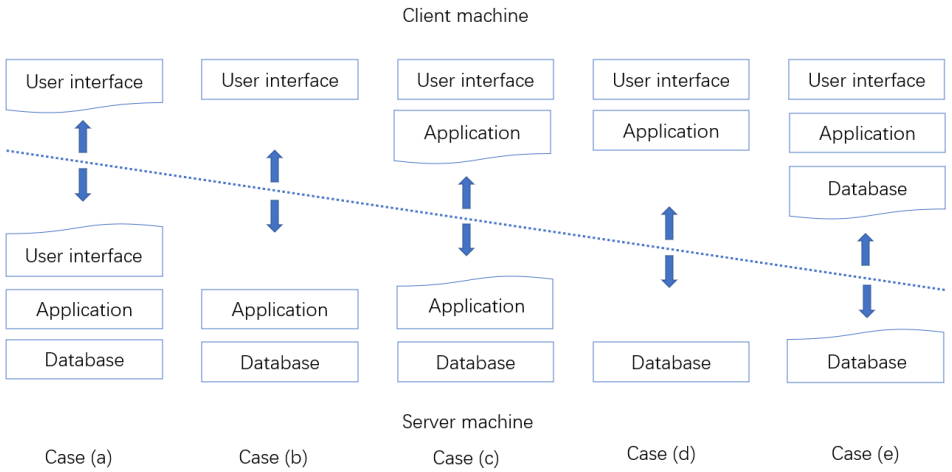


Figure 1. Alternative client-server organizations

Video object detection program design under different client-server organizations

Based on alternative client-server organization, video object detection program will be design under Case (b) and Case (c) respectively (shown in Figure 1). For Case (b), since the client only provides the user interface, the main function of the user interface is to upload videos, and the processing logic related to video processing and object detection is completely handed over to the server. For Case (c), since the client not only provides the user interface, but also provides some application processing logic, the client is mainly responsible for uploading video and video processing logic, and the processing related to object detection is handed over to the server.

1. Video object detection program design under Case (b)

Based on client-server organization Case (b), all processing logic is placed on the server side, and the client only provides interfaces for uploading and playing videos. The video object detection program under Case (b) is designed in the scenario where the user accesses the browser. Figure 2 shows the interactive process of realizing the video object detection function through a sequence diagram. In Figure 2, Browser represents User interface that provides the function of uploading video and playing video. Server and Live Video Server represent Application that provides all processing logic for video object detection, including video segmentation, object detection of frames, real-time transmission of frames by streaming. Database mainly records the processed video information, such as processing time, frame size, frame rate, etc.

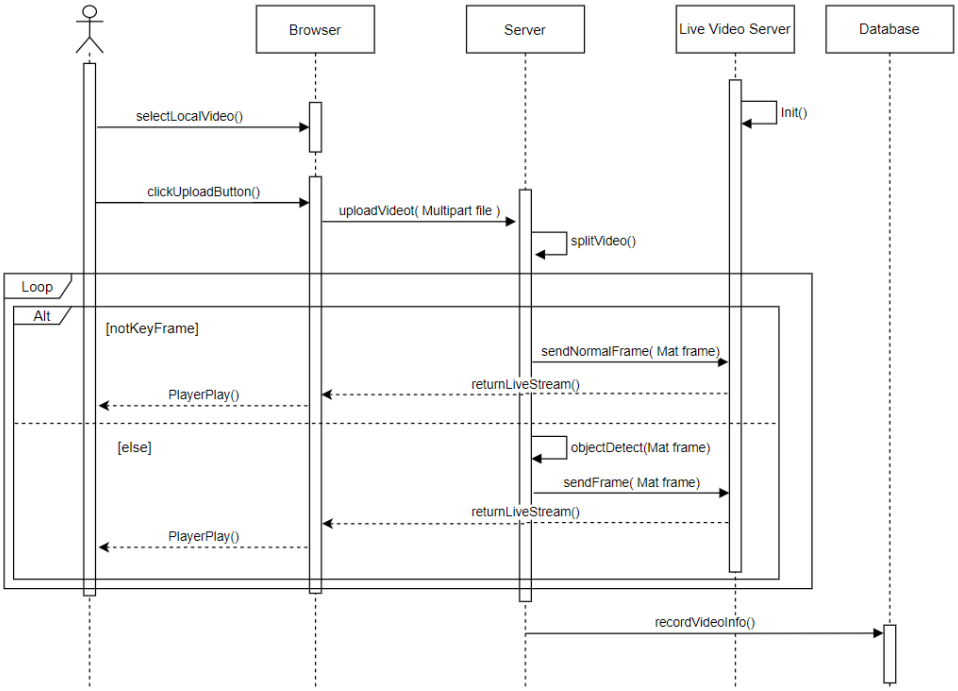


Figure 2. Sequence diagram of Case (b)

2. Video object detection program design under Case (c)

Based on client-server organization Case (c), the client side mainly includes interfaces for uploading and playing videos, video processing logic and frame processing logic, while the server side mainly handles frame object detection. The video object detection program under Case (c) is designed for mobile devices. Users can access the server through mobile devices to request image object detection, and cooperate with the processing logic on the mobile device to complete video object detection. Figure 3 shows the interactive process of realizing the video object detection function through a sequence diagram. In Figure 3, APP represents User interface and Application of client side that provides the functions of local video selection, video segmentation, key frame selection, image processing, composite video and video playback. Server represent Application of server side that provides image object detection. Database mainly records the processed image information, such as processing time, image size, etc.

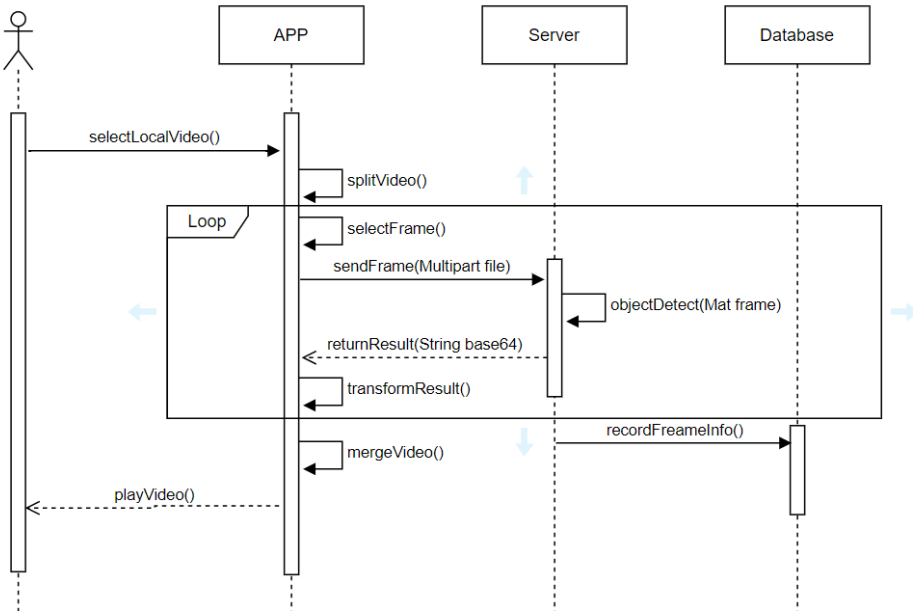


Figure 3. Sequence diagram of Case (c)

3. The process of image object detection

In the process of video object detection program design, it involves object detection of images (object detection of frames). The image object detection method based on YOLOX mainly involves image preprocessing, output decoding and the strategy for choosing prediction boxes. The process of image object detection mainly includes the following steps:

Step 1. Image transformation and normalization.

The purpose of transforming and normalizing the images is to obtain images that satisfies the input format of the YOLOX model. The YOLOX model requires the size of the input image to be 640×640 pixels. According to the affine transformation theory, image transformation can be done with the help of transformation matrix M. For image normalization, using the mean and std of ImageNet is a common practice. All models expect input images normalized in the same way, mini-batches of 3-channel RGB images of shape (3×H×W), where H and W are expected to be at least 224. The images have to be loaded in to a range of [0, 1] and then normalized using mean with [0,485; 0,456;0,406] and std with [0,229; 0,224; 0,225].

Step 2. Decode outputs, generate proposals.

The output of the model is the tensor of 8400×85 elements, meaning that the model predicts 8400 prediction boxes and every prediction box has 85 properties. Among the 85 properties, the first five properties represent the horizontal and vertical coordinates of the center point of the prediction box, the length and width of the prediction box, and the probability of objects in the prediction box. The remaining 80 attributes represent the object category probability (the model can judge 80 object categories, and each object category will correspond to a probability).

For an image, there are often only a few objects that need to be identified. However, there are 8400 boxes based on the output data. Obviously, the prediction boxes need to be further selected. In this step, it is necessary to manually set threshold to generate proposals with more accurate prediction boxes.

Before manually setting the threshold, it is necessary to know that the output 8400 prediction boxes are merged from three different sizes of anchors. Because in the decoding process, the coordinates of the prediction boxes are related to the corresponding anchor. Among the 8400 prediction boxes, the anchor size corresponding to 6400 prediction boxes is 8×8. This means that the original image of size 640×640 is divided into 80×80 anchors with stride of 8. Similarly, the anchor size corresponding to 1600 prediction boxes is 16×16, and the anchor size corresponding to 400 prediction boxes is 32×32.

Through the anchor, the coordinates of the prediction box can be obtained, and the probability threshold is further set manually. Here the probability threshold is set to 0,6. Then the confidence C that each prediction box has the corresponding category of objects can be given

$$C = p \cdot \max(class\ probability) \tag{5}$$

where p represents the probability that there is an object in the predicted box. Class probability represents the probability of being that class. Then if the confidence C is greater than 0,6, save the proposal (the coordinates of the prediction box) and the corresponding the confidence C. After generating the proposals, the possibly correct prediction boxes will be saved.

Step 3. Perform non-maximum suppression.

The purpose of performing the non-maximum suppression algorithm is to eliminate redundant overlapping boxes with lower confidences.

Step 4. Accomplish object detection.

According to the affine transformation theory, based on the transformation matrix M^{-1} from Image transformation, the coordinates of the final boxes can be converted to the coordinates in the original image. It accomplishes the object detection on the original image.

Test Result

The developed method of image object detection is implemented in C++ using library of OpenCV 4.5.4 and the part of splitting video is implemented using library of FFmpeg. For the program designed based on Case (b), the front-end program is developed based on Vue3 and deployed on Amazon Elastic Compute Cloud (EC2) together with the back-end server. For the program designed based on Case (c), the front-end application is developed on the Android system, and the back-end server is

deployed on EC2. EC2 is physically located in Frankfurt. The EC2 has the following specifications: vCPU: 1 core; Memory: 1GB; for bandwidth, the available network bandwidth of an instance depends on the number of vCPUs it has.

Examples of test result are shown in Figure 4 and Figure 5.

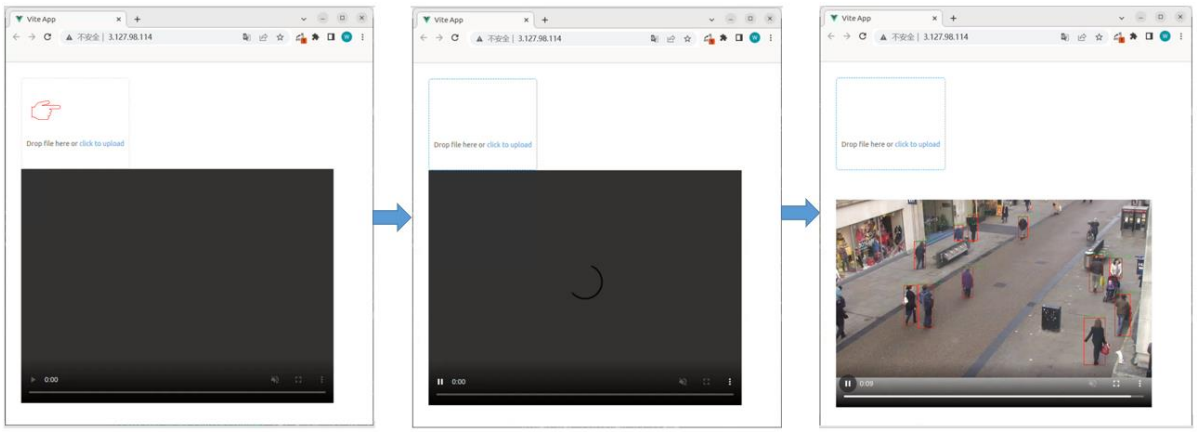


Figure 4. Test Result of Case (b)

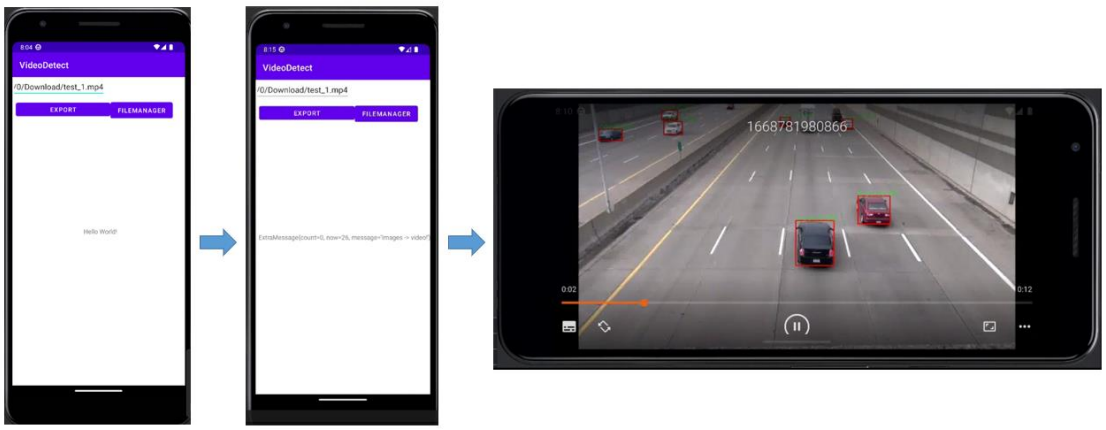


Figure 5. Test Result of Case (c)

In Figure 4, video object detection program is designed based on client-server organization Case (b). The user can open the browser to upload the video, wait for the buffering of the live stream, and after some data is processed in the background, the user can watch the result of video object detection. In Figure 5, video object detection program is designed based on client-server organization Case (c). Through the application on the mobile phone, the user can select a local video, or input the path of the video, and then clicks the button. The application starts to process the video and interacts with the background server to complete object detection, and finally synthesizes the video and plays it.

For Case (b), two sets of tests are carried out. The test object metadata is given in Table 1 and the test data are shown in the Table 2.

Table 1. **Metadata**

Video ID	Resolution, pixel	Frame Rate, FPS	BPS, kb/s	Size, Mb
1	1280×720	30	9818	39
2	1280×720	30	4321	17

Table 2. **The test data**

Video ID	Upload Time, s	Wait Response Time, s	Steam Content, Mb	Transmission Time, s
1	30,56	1,63	10,1	174
2	7,46	1,43	11,4	168

In table 2, Upload Time represents the upload time from client to server. Wait Response Time represents the time from when the server returns the message of receiving the uploaded video to when it starts to receive streaming data. Steam Content represents the size of the stream data received. Transmission Time represents the time to receive streaming data. For Upload Time, it depends on the

local upstream bandwidth. If the upstream bandwidth is smaller, the upload time will be longer. For Wait Response Time, during this period, the server needs to pre-process the first few frames, and then start streaming. For Steam Content and Transmission Time, they not only depend on the local bandwidth, but also depend on the upstream bandwidth of the server. Because the local bandwidth is 10M, the general network speed is 1,25 Mb/s. However, the test data does not meet expectations. The main reason is that the actual uplink speed of the server is too low. The part of real-time monitoring data of the server network speed is shown in the Figure 6.

Time	lo		eth0	
HH:MM:SS	KB/s in	KB/s out	KB/s in	KB/s out
19:01:36	0.00	0.00	105.46	204.65
19:01:37	0.00	0.00	24.72	47.45
19:01:38	0.00	0.00	86.74	168.45
19:01:39	0.00	0.00	56.77	109.53
19:01:40	0.00	0.00	28.02	54.51
19:01:41	0.00	0.00	61.47	118.79
19:01:42	0.00	0.00	65.35	126.04
19:01:43	0.00	0.00	52.00	99.58
19:01:44	0.00	0.00	50.13	97.26
19:01:45	0.00	0.00	81.20	156.07
19:01:46	0.00	0.00	7.95	16.15
19:01:47	0.00	0.00	57.97	111.48
19:01:48	0.00	0.00	97.50	188.95
19:01:49	0.00	0.00	17.52	34.01
19:01:50	0.00	0.00	99.33	192.20
19:01:51	0.00	0.00	68.11	132.00
19:01:52	0.00	0.00	41.64	80.95
19:01:53	0.00	0.00	113.67	219.96
19:01:54	0.00	0.00	38.09	73.07
19:01:55	0.00	0.00	64.28	125.32
19:01:56	0.00	0.00	101.81	196.64

Time	lo		eth0	
HH:MM:SS	KB/s in	KB/s out	KB/s in	KB/s out
19:03:00	0.00	0.00	52.31	101.05
19:03:01	0.00	0.00	48.93	94.57
19:03:02	0.00	0.00	56.64	109.04
19:03:03	0.00	0.00	73.63	142.28
19:03:04	0.00	0.00	26.86	52.89
19:03:05	0.00	0.00	32.48	61.57
19:03:06	0.00	0.00	142.48	276.33
19:03:07	0.00	0.00	70.49	136.40
19:03:08	0.00	0.00	32.13	62.74
19:03:09	0.00	0.00	110.47	214.26
19:03:10	0.00	0.00	52.47	102.53
19:03:11	0.00	0.00	58.75	112.22
19:03:12	0.00	0.00	74.14	143.50
19:03:13	0.00	0.00	29.33	57.85
19:03:14	0.00	0.00	59.72	114.31
19:03:15	0.00	0.00	114.18	221.43
19:03:16	0.00	0.00	26.20	50.65
19:03:17	0.00	0.00	110.90	215.15
19:03:18	0.00	0.00	76.73	147.93
19:03:19	0.00	0.00	27.85	54.99
19:03:20	0.00	0.00	120.19	233.06

Figure 6. Server real-time network speed monitoring data

In Figure 6, eth0 represents the network card of the server, and "in" and "out" represents download speed and uplink speed respectively. It can be seen from the Figure 6 that the uplink speed of the service is too low and unstable, which directly causes the client to freeze or drop frames when playing video. This degrades the experience of watching live.

For Case (b), the upload network speed of the client and the download network speed of the server determine the start time of the user waiting for the live broadcast. The download speed of the client and the upload speed of the server determine the user experience of watching the object detection video. For users, 10M bandwidth is enough to upload video and watch live broadcast. However, if the upstream bandwidth of the server is not large enough, it will directly cause the live video to freeze or drop frames. In this case, Case (b) is not recommended. For Case (c), the network transmission part only involves the transmission of images. The requirement for server bandwidth is not as high as Case (b). Therefore, when the bandwidth of the client and the bandwidth of the server cannot be guaranteed, Case (c) can be selected. In this case, the waiting time may be too long, but this case is more general with less material hardware and bandwidth requirements.

Conclusion

Under two client-server organization cases, two designs of video object detection programs are proposed. One is to put all the processing logic of video object detection on the server side, and the client only provides the interface; the other is to put only the function of object detection on the server side, and other video processing logic on the client side. Then two designs are implemented and tested by building a cloud server. The result is that the proposed program design can complete video object detection. In addition, the use conditions of the two designs are analyzed through the test data.

References

1. Jing J, Helal A. S., Elmagarmid A. // ACM Computing Surveys. 1999. Vol. 31. P. 117-157.
2. Redmon J., Divvala S., Girshick R., et al. You Only Look Once: Unified, Real-Time Object Detection. J. 2016.
3. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. J. 2018.
4. Ge Z., Liu S., Wang F., et al. YOLOX: Exceeding YOLO Series in 2021. J. 2021.
5. Neubeck A., Gool L. Efficient Non-Maximum Suppression. C. 2006.
6. Stearns C., Kannappan K. Method for 2-D affine transformation of images. P. 1995.

УДК 330.344.24

ПОДХОД К ПОСТРОЕНИЮ ПОДСИСТЕМ УМНОГО ГОРОДА

В.А. ВИШНЯКОВ, В.А. ГРОМОВ, С.В. КУЧЕРОВ,
С.А. СИДОРЕНКО, А.В. УСЕВИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Представлена концепция умного города (УГ). Рассмотрены подсистемы УГ энергетики, транспорта, логистики, электронного правительства. Обсуждены их структуры и функции. Определены направления по разработке данных подсистем на базе платформы Интернет вещей.

Ключевые слова: умный город, энергетика, транспорт, логистика, электронное правительство.

Введение

Быстро развивающиеся интеллектуальные информационно-коммуникационные технологии (ИКТ) трансформируют все сферы, в том числе традиционные методы работы органов власти. Граждане ожидают от государства более совершенных электронных форматов предоставления государственных услуг, с одной стороны, и улучшения форм коммуникации с целью развития прямой демократии, с другой стороны. Это обеспечивает система Умного города.

Понятие «Умный город» (УГ) можно определить как «…применение информационно-коммуникационных технологий с их воздействием на человеческий капитал/образование, социальный и реляционный капитал и экологические проблемы». Приведем и другие определения понятия УГ, концепция которого больше не ограничивается ИКТ, а рассматривает потребности людей и сообщества. «Город, который контролирует и интегрирует работу всех своих критических инфраструктур, включая дороги, мосты, туннели, рельсы, метро, аэропорты, морские порты, коммуникации, водо- и электроснабжение, даже крупные здания, который умеет оптимизировать свои ресурсы, планировать профилактические мероприятия и контролировать безопасность, при этом максимизируя полезность услуг для своих граждан» [1].

С точки зрения технологического подхода [2] умный город – это город с большим присутствием ИКТ нового поколения, применяемых к критически важным компонентам городской инфраструктуры и услугам: наличие умных систем управления дорожным движением; умный подход к уличному освещению; внедрение общегородской и доступной сети Wi-Fi; использование умных сетей и альтернативных источников энергии; наличие системы оповещения граждан о чрезвычайных ситуациях; минимальное использование наличных средств для оплаты товаров и услуг; активное привлечение граждан к вопросам городского управления. В состав системы УГ входят более десятка подсистем, рассмотрим четыре из них, соответствующие критически важным компонентам.

Подсистема УГ энергетика

Умная энергетика нужна как государству, так и министерству энергетики, в первую очередь она поможет выйти на новый уровень автоматизации, а также производить мониторинг

и анализировать данные. Государство по результатам мониторинга может вносить изменения на законодательном уровне в сфере энергетики и давать стратегические посылы для развития отрасли. Энергетика относится к стратегическим областям, и государство напрямую заинтересованно в ее унификации и оптимизации. Помимо мониторинга подсистема позволит: увеличить рост доходов, за счет оптимизации электроснабжения; снизить потери на ЛЭП; экономить ресурсы; ускорить сроки планирования нагрузок сетей; уменьшить время замены вышедшего из строя оборудования, за счет датчиков, отслеживающих жизненный цикл данного оборудования (вовремя формировать заявку на поставку запчастей, нового оборудования).

Эти технологии актуальны в Республике Беларусь и России, которые обладает огромной централизованной системой энергоснабжения. На уровне управления системой, балансами и режимами в электроэнергетике, интернет вещей позволит более рационально планировать загрузку генерирующих мощностей и их объем [3]. Что касается электросетевого хозяйства, то индустриальный интернет вещей позволил бы (с учетом протяженности ЛЭП) повысит надежность и снизить операционные расходы. Появляется возможность производить ремонтные работы сети «по состоянию» электросети, а не по регламенту обслуживания, что гораздо снизит затраты на ремонтные работы. Можно сделать следующие выводы по использованию подсистемы энергетики:

1. Интернет вещей (IoT) позволит оптимизировать энергоснабжение, уменьшить потери электроэнергии, производить мониторинги и модернизировать систему электроснабжения.
2. В Республике Беларусь есть огромные перспективы для развития индустриального интернета в энергетике, так как многие проекты уже запущены в пилотном режиме.
3. Сложность перехода заключается в дорогой стоимости оборудования, и главным образом, в консервативности сферы энергетики, которая обеспечивает прежде всего безопасность и надежность.

Концептуально IoT может применяться (и применяется) в сфере энергетики в двух глобальных направлениях – снижение потребления ресурсов и отслеживание технического состояния оборудования с целью проведения своевременного технического обслуживания, а также предупреждения аварийных ситуаций. В энергетике сенсоры и датчики, подключенные к Интернету, используются для построения «умных» электросетей и инфраструктуры (Smart Grids).

Подсистема УГ транспорт

Подсистема транспорта «Умного города» основывается на интеллектуальной транспортной системе (ИТС). Функции ИТС обеспечивают интеграцию оперативного управления всеми видами городского транспорта и возможность реакции на события в режиме реального времени [4]. Главная инновация «Умного города» в отношении транспорта – это создание города, ориентированного на пешехода и стремление свести использование частного транспорта к минимуму. Поэтому серьезное внимание в транспортной системе уделяется общественному транспорту, его доступности, информированности о расписании, электронная оплата и т.д.

Критичные для успешного функционирования системы узлы – это в первую очередь транспортно-пересадочные пункты, куда входят также перехватывающие паркинги. Для того чтобы обеспечить их функционирование, необходима интеграция информационных и навигационных систем в рамках единой платформы «Умного города». Большое значение в ИТС имеет наличие единого транспортного интерфейса, ориентированного на потребности жителей УГ и гостей, внутри которого можно найти и использовать множество сервисов – от подсказки, на какую парковку вести машину, до оповещения о сроке прибытия местного общественного транспорта [4].

В городе должен быть создан единый центр управления ИТС, куда будут в онлайн-режиме передаваться данные с детекторов мониторинга транспортных потоков и дорожная обстановка с фото- и видеокамер. Система также должна фиксировать скорость потока, количество автомобилей и общественного транспорта, метеоусловия и состояние трассы. В случае ДТП система должна предупреждать о затруднениях на дороге и подсказывать объездные пути. Сигналы светофоров должны меняться в зависимости от загруженности соседних перекрестков.

При этом появится возможность координировать потоки в случае заторов, отменять непопулярные маршруты и назначать новые.

Умным принято называть светофор, которым управляет специальная программа, позволяющая устройству самостоятельно принимать решения, в том числе на основе поступающей информации о дорожном движении с других аналогичных приборов.

Дорожные камеры выступают «глазами» современных интеллектуальных транспортных систем. Это камеры высокого разрешения, которые повсеместно используются разработчиками ИТС и комплексов видеофиксации нарушений ПДД. Информационные табло – это основное средство информирования водителей о ситуации на дорогах. На табло может выводиться различная информация: загрузка участков дороги; наличие ДТП на маршруте; количество общественного транспорта; состояние дорог и т.д.

Подсистема УГ логистика

Основным из аспектов подсистемы «Умный город» является реализации логистических цепочек в условиях становления современного общества. Для реализации данной концепции логистика подсистемы «Умный город» требуется сочетание в себе организационных инноваций с инфокоммуникационными технологиями для выполнения цифровизации городов [5]. Идея применения данной подсистемы в том, что высокая эффективность реализации проектов логистики приводит к росту уровня жизни городов. На основании отечественного опыта так и опыта различных городов в странах мира, проведен анализ реализации задач умной логистики.

В Российской федерации преобладают автомобильные перевозки. Для современного мира ключевыми параметрами в логистической отрасли является быстрота доставки товара, а также стоимость доставки. Для рынка грузоперевозок требуются инновации, которые будут оптимизировать именно вышеуказанные параметры. Оптимизация доставки товаров положительно скажется на всем обществе, позволив разгрузить транспортные магистрали и снизить трафик в том числе и внутри городов России, также повсеместное использование концепции смарт логистики позволит снизить как время, так и стоимость доставки грузов, что тоже немаловажный показатель в современных экономических отношениях [6].

Для эффективной реализации подсистемы логистика УГ необходимо совершенствование механизма информационного обеспечения, реструктуризация и модернизация государственных органов. Подсистемы транспорта и логистики не может работать отдельно и требуется полное сотрудничество со специалистами телекоммуникационных, навигационных и информационных технологий.

Подсистема УГ электронное правительство

Термин «Smart Government» довольно часто употребляется в научных дискуссиях и исследованиях масштабного феномена «Smart Society», объясняющего социально-экономическую, политическую и культурную стратегию развития общества на основе обширной цифровизации всех сфер жизни. В более узком смысле междисциплинарных учения явления «Smart City» также встречается использование термина «Smart Government». Вместе с тем в этой области термин «смарт-правительство» зачастую является просто синонимом сходных определений: «электронное правительство» (e-Government) и «электронное управление» (e-Governance).

Смарт-правительство – это развитое электронное правительство, сформированное на открытом управлении, которое использует преимущества, предоставляемые ИКТ, собирая и обобщая физическую, цифровую, государственную и частную среды для пассивного и активного взаимодействия и сотрудничества с гражданами с целью улучшения понимания их потребностей и творческого, эффективного и гибкого оказания услуг в любом месте и в любое время, в том числе предикативно [7].

Платформа Интернет вещей для реализации подсистем УГ

Первая очередь платформы УГ обеспечивает: приложения для деятельности местных органов власти и поддержки принятия ими управленческих решений (контроль выполнения поручений, отслеживание ключевых показателей успеха, нормативно-справочная информация); сервисы электронного участия (заявки, работники, общественное мнение); городские сервисы (городская информация, вакансии, торговые площадки), туристические сервисы (исторические объекты, экскурсии, навигация); деловые сервисы.

Система CitySys представляет собой открытую платформу, объединяющую в себе множество приложений для организации умного города. Сбор данных, передача и оценка выполняются комплексной системой управления CitySys, реализованной на платформе ThingsBoard IoT в рамках стандарта открытого исходного кода (OPC) [8].

Информационное взаимодействие платформы с открытым исходным кодом – это серия спецификаций от поставщиков и разработчиков программного обеспечения, определяющих интерфейс между клиентами и серверами, включая доступ к данным в реальном времени, мониторинг аварийных сигналов и событий, доступ к историческим данным и другие области применения. Его аппаратные компоненты обеспечивают прямую связь через стандартные интерфейсы и протоколы, например: Powerline, Bluetooth, KNX, Z-Wawe, ModBus RTU/TCP, BACnet IP, EnOcean, DMX, M-Bus, GSM, 1-wire и DALI. Кроме того, предлагается выдача стандартизованного API, в частности REST API, d. Для связи между фонарями нами используется передача по линиям электросети. Это означает, что коммуникационный сигнал передается по стандартной электрической сети напряжением 230 В. Учитывая подключение к сторонним системам и то, что в городе уже установлены существующие системы, Citysys открыта для коммуникационных протоколов MQTT, JSON, XML, XMPP, SMTP и RSS. Накопленные данные хранятся на облачном сервере [8].

IoT платформа Умный город CitySys является горизонтально масштабируемой и строится с использованием технологий с открытым исходным кодом. Благодаря идентичности каждого узла кластера, платформа становится полностью отказоустойчивой. Надежность и высокая эффективность позволяют одному узлу управлять десятью или даже сотнями тысяч устройств. Платформа IoT умного города содержит настраиваемые виджеты, механизм обработки правил и систему подключаемых модулей, которые фактически делают платформу расширяемой. Сбор данных в умном городе обеспечивается множеством датчиков, отслеживающих различные параметры: датчики движения, датчики дорожного движения, детекторы заполнения парковочных мест, погодные станции, датчики управления отходами, датчики шума, обнаружения стрельбы, камеры CCTV и кнопки экстренного вызова [8].

Заключение

Обсуждена концепция и структура умного города. Описаны оссобенности четырех подсистем УГ: электроэнергетики, транспорта логистики и электронного правительства. Рассмотрены особенности IoT платформа Умный город CitySys для реализации подсистем.

AN APPROACH TO THE CONSTRUCTION OF SMART CITY SUBSYSTEMS

U.A. VISHNYAKOU, V.A. GROMOV, S.V. KUCHEROV,
S.A. SIDORENKO, A.V. USEVICH

Abstract. The concept of a smart city (SC) is presented. The subsystems of the SC of energy, transport, logistics, and e-government are considered. Their structures and functions are discussed. The directions for the development of these subsystems based on the Internet of Things platform are defined.

Keywords: smart city, energy, transport, logistics, e-government.

Список литературы

1. Albino V., Berardi U., Dangelico R.M. // Journal of Urban Technology. 2015. Vol. 22(1). P. 1–18.
2. Головенчиков Г.Г., Цзяньвэ С. // Вестник сувязи. 2023. № 1. С. 40–45.
3. Smart Grid или умные сети электроснабжения. [Электронный ресурс]. URL: https://eneca.by/novosti/energetika-i-energoeffektivnost/smart-grid-ili-umnye-seti-elektrosnabzheniya.
4. Кондукова Ю.Ю. // Молодой ученый. 2021. № 50 (392). С. 66–69.
5. Интеллектуальные Логистические Технологии. [Электронный ресурс]. URL: https://www.hitachi-transportsystem.com/en/solution/smartlogistics/.
6. Сыров М.С. и др. // Российские регионы в фокусе перемен. Т. 2. Екатеринбург, УрФУ. 2022. С. 332–335.
7. Bradul N.V., Lebezova E.M. // Upravlenets – The Manager. 2020. Vol. 11(3). P. 33–45.
8. IoT платформа Умный город CitySys. [Электронный ресурс]. URL: https://www.intelvision.ru/products/platforma-umnyi-gorod.

UDC 621.391

ANALYSIS AND SIMULATION BASED ON 5G CHANNEL CODING TECHNOLOGY

F.Y. LIU, S.B. SALOMATIN, K.V. MIHNO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 10, 2023

Abstract. In wireless communication systems, the noise and electromagnetic interference in the channel can have a significant impact on the transmitted signal. The introduction of channel coding techniques is a very effective way to reduce the bit-error-rate (BER). As mobile communication has entered the 5G era, the data traffic has increased significantly compared to 4G, and as the modulation steps increase, the BER will increase, which makes the importance of channel coding even more significant. In China Mobile's 5G technology, channel coding is usually performed by LDPC (Low Density Parity Check Code), and the application of channel coding technology in China Mobile's 5G technology and how to improve the coding rate will be discussed.

Keywords: 5G channel coding, BER, LDPC.

Introduction

5G is a new generation of wireless mobile communication network that is currently being vigorously studied by China Mobile, and the 100-megabit data transmission will bring unprecedented pressure to the entire network. The reason why channel coding technology is introduced in the digital communication system is that the noise in the channel interferes with the transmission signal and the error code often appears in the received signal. In an environment of extremely high growth of transmitted data, how to ensure a very low bit error rate will be a new challenge for 5G networks. Channel coding is a way to effectively control the BER using forward error correction by adding redundancy to the transmit signal while eliminating these redundant codes in the signal receiver using a decoder. These redundant codes enhance the confidence level of the signal and minimize other effects on the signal such as noise, thus reducing the BER. In the 1960s, LDPC codes, low-density parity-check codes, were first proposed by Dr. Gallager, but limited by the technical conditions at that time, there was a lack of feasible decoding algorithms until the 1990s, when a breakthrough was made based on Turbo codes.

LDPC code, which borrows the circular iteration mechanism from Turbo code, is a linear code generated from a recursive convolutional encoder that uses an iterative decoding process to decode the received information. The recursive convolutional encoder is a group code that converts bits of information into 1 group length, and each group code containing transmitted information can be decoded into the original information. In an iterative receiver, the decoder usually shows the probability of receiving a digital signal "1" or "0", and the probability of receiving a signal can increase with the number of iterations. The iterative receiver is composed of two identical decoders, each of which can also use the result of the other to produce a more accurate signal probability, which is the whole process of iterative decoding. In the channel coding parallel system, the transmitter side usually consists of two encoders and mixers, while the receiver side consists of two decoders and interleavers, and the two structures complement each other B In the encoder part, China Mobile 5G [1–5] system uses convolutional codes for coding, and this paper will also explore LDPC recursive system convolution. This convolutional code can correct transmission errors by adding redundant codes to the transmit information, while helping the decoder to improve the trustworthiness of the received signal. Therefore, adding convolutional codes in communication systems is to convert the transmitted bits into a longer set

of strings, and the reason why longer information is used to transmit is to correct the effect of channel noise on the transmitted information. Finally, the final received signal is obtained through the signal "trustworthiness" after several iterations, and when the number of iterations increases, the false bit rate decreases, thus completing the whole channel coding process.

Introduction of LDPC coding technology

LDPC codes, which are mapped from message sequences into sending sequences, code word sequences, by passing a generation matrix *G*. For the generation matrix G, there exists a fully equivalent parity check matrix ugly, and all the code word sequences V constitute the zero space of H, the number of non-zero elements in each row and column of the $H \cdot V = 0$ check matrix twins is very small, which is the reason why LDPC codes are called low density codes. The message to be sent is mapped into a transmit sequence by a specific checksum matrix and decoded in the decoder using the circular iteration principle, and the log-likelihood ratio of the channel output y with respect to the transmitted bits a is calculated as equation (1) below, where L_c – channel confidence.

$$A_c(a) = \lg \frac{p(y \mid a = 0)}{p(y \mid a = 1)} = L_c y. \tag{6}$$

Figure 1 shows the channel coding schematic, which is widely utilized in existing communication systems. In order to show the channel coding system more systematically and concretely, this paper proposes to use QPSK debugging as the debugging signal, the encoder will utilize LDPC convolutional codes, the interleaver is the most common hybrid coding; the receiver is the corresponding QPSK demodulator, and the de-interleaver. The interleaver and the deinterleaver have opposite functions, in order to disrupt the order of the transmitted bits and thus reduce the adjacent bit interference to a lower level. The channel noise is the most common Gaussian white noise.

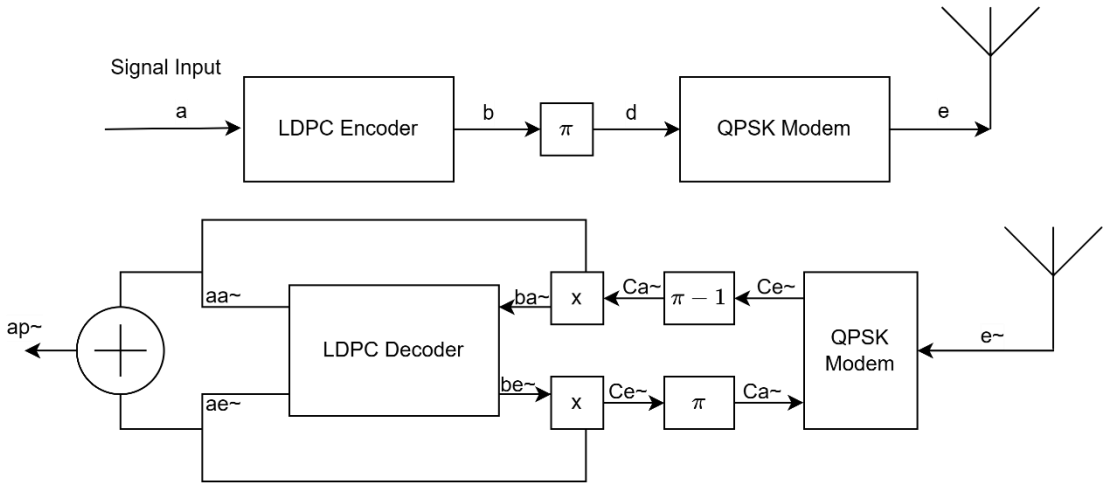


Figure 2. Schematic diagram of LDPC recursive convolutional coding system

LDPC-Conv codes can be regarded as a generalization of LDPC codes in the time domain, which can provide a certain degree of coding gain. The decoding algorithm of LDPC-Conv codes is also based on a probabilistic transfer algorithm. Specifically, the decoding algorithm of LDPC-Conv codes can be regarded as a convolutional version of LDPC codes, and its decoding algorithm is similar to that of LDPC codes, which is also an iterative algorithm based on message passing.

The iterative algorithm is a computational method that approximates the solution to a problem systematically through a series of repeated computations. It is an algorithm that decomposes a problem into small subproblems, solves each subproblem individually, and combines the solutions of the subproblems to obtain a solution to the original problem. Iterative algorithms are particularly suitable for situations where the problem is too complex to be solved by analytical methods, or where an exact

solution is not available. In such cases, iterative algorithms provide a useful approximation to the solution.

The transmitted bits pass through the convolutional encoder and the interleaver in the transmitter, respectively, and then modulated into the channel, this time the transmitted information is a coded QPSK signal with Gaussian white noise; at the receiver side, the demodulator will generate a possibility of receiving the information (as "1" or "0"), and the received signal will be iterated between the interleaver and the decoder demodulator. Through the iterative algorithm, each demodulation signal C_e is based on the received signal $\hat{e}$ as well as the previous received signal C_a. When several iterations are completed, the system can judge by itself whether the current $ap\sim$ is the best output and the output result is not after optimization, then $ap\sim$ is the last received signal, so the final received signal will be the result after several iterations of error correction. In the LDPC channel coding system, iterative decoding will be used for both decoder and demodulator. In each iterative decoding process, the accuracy of output bits will increase with the end of each iteration process, and after several iterations, the system can output the best result. The number of iterations can be determined by the parameters of the system noise and the signal-to-noise ratio (SNR). The channel coding system with iterative decoding can approach the theoretical limit of the Shannon capability, which is the main reason for introducing the channel coding technique.

The top line shows the theoretical bit error rate plot for the QPSK system, i.e., the system without channel coding, and the middle and bottom lines show the bit error rate plots for the LDPC channel coding technique after 1 iteration and 8 iterations, respectively.

Simulation of cyclic iterative coding

In the above discussion, the LDPC decoder is a buffered decoder, and only one data can be processed in each iteration, because the decoder or demodulator cannot work at the same time, and the demodulator or decoder needs to wait for the data from the other side during the loop iteration, so there is always an "idle" part in the whole system. The purpose of bufferless decoder is to eliminate this idle state, so that the decoder and demodulator can work at the same time and get the received signal faster.

The decoder can work at any time as long as the receiver is working. The use of cacheless decoders will greatly improve the decoding efficiency, and it has been found through extensive experiments that cacheless decoders can use fewer iterations to complete demodulation than conventional demodulators for the same channel, due to the fact that the decoding process and demodulation process are carried out simultaneously, so that the best results can be achieved earlier.

The comparison of the simulation results in Figure 2 shows that the cacheless decoder stops working after two iterations, while the regular decoder performs eight iterations and obtains essentially the same results.

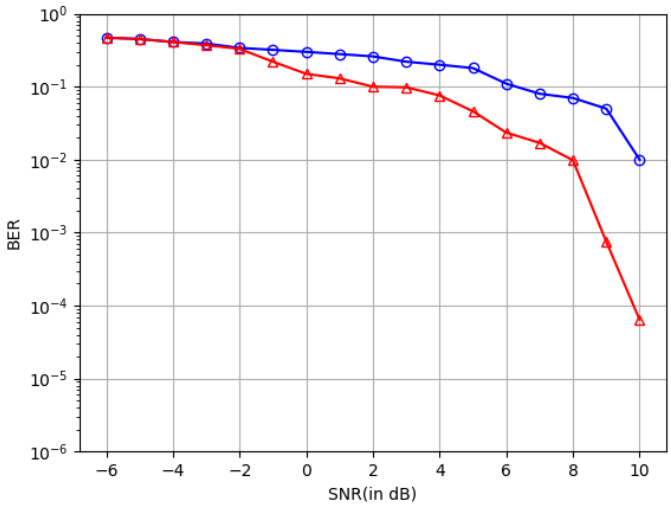


Figure 3. Comparison of the simulation results of LDP C cacheless decoders and cacheless decoders: Red – cacheless decoders; Blue – cacheless decoders

Conclusion

The introduction of channel coding is effective in reducing the false bit rate in the context of the increasingly aggressive electromagnetic environment and the large capacity transmission of 5G by China Mobile. LDPC's loop iterative coding can be effectively applied to 5G communication networks. According to the BER graphs obtained for different number of iterations, the lower iteration system requires higher SNR for the same FER, while the higher iteration system has lower BER for the same SNR. However, when the number of iterations is increased to a certain number, the BER of the system does not decrease indefinitely, because the system is already close to the Shannon Limit Theory. In the decoder part, the conventional decoder can be optimized to a cacheless decoder, which can obtain almost the same result faster under the same conditions, greatly improving the efficiency of channel coding, and will have a profound impact on the future development of 5G channel coding.

References

1. Darabiha A., Carusone E. // A bit-seral approximate min-sum LDPC decoder and FPGA implementation. 2006. P. 149–152.
2. Gunnam K., Catala Perez M. // Algorithms and VLSI architectures for low-density parity-check codes. 2016. P. 57–63.
3. Elidan G., Koller D. // Conference on Uncertainty in Artificial Inteligence. 2006. P. 165–173.
4. Vila Casado A., Griot M., Wesel D. // IEEE Transactions on Communications. 2010. P. 3470–3479.
5. Richardson J., Shokrollahi M. // IEEE Transactions on Informaation Theory.2001. P. 619–637.

УДК 004.932+004.258

СКЕЛЕТИРОВАНИЕ НИЗКОКОНТРАСТНЫХ ЗАШУМНЫХ СЕРЫХ ИЗОБРАЖЕНИЙ

Ц. МА, А.А. БОРИСКЕВИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 13 Марта 2023

Аннотация. Для повышения устойчивости скелетов полутоновых изображений с двухмодовой гистограммой яркости к шуму в статье предложена модель скелетизации, учитывающая наличие мультипликативной и аддитивной составляющих шума на бинарном скелетизируемом изображении. С учетом данной модели разработан алгоритм скелетизации, отличающийся учетом искажений форм областей скелетизируемого бинарного изображения в результате низкочастотной фильтрации исходного полутонового изображения и позволяющий уменьшить ошибки скелетизации полутоновых изображений.

Ключевые слова: скелетизация полутоновых изображений, мультипликативный шум, аддитивный шум, чувствительность скелета к шуму.

Введение

Алгоритмы скелетизации полутоновых изображений формируют более устойчивые к шуму скелеты. Они основаны на предварительной низкочастотной фильтрации исходных полутоновых изображений с использованием неориентированных и ориентированных в пространстве Гаусс-фильтров в сочетании с поиском в разномасштабных версиях исходного изображения значимых точек или однородных по яркости областей, подбором параметров фильтра, обеспечивающих наименьшую чувствительность скелета к шуму. Среди этих алгоритмов наиболее эффективен алгоритм, в котором выбор параметров низкочастотного фильтра основан на вычислении минимального значения метрики чувствительности скелета к шуму [1–4].

Скелет может формироваться с помощью любого алгоритма бинарной скелетизации. Недостатки алгоритма [2] состоят в том, что: а) используемая метрика чувствительности скелета к шуму не учитывает искажения форм областей скелетизируемого бинарного изображения в результате низкочастотной фильтрации исходного полутонового изображения; б) не определены значения отношения «сигнал/шум» для исходного полутонового изображения при которых полутоновая скелетизация позволяет уменьшить ошибки формирования скелетов по сравнению с бинарной скелетизацией.

Целью работы являются уменьшение ошибок скелетизации полутоновых изображений с двухмодовым распределением яркости в условиях шума и определение значений отношения «сигнал/шум» для эффективного использования скелетизации на основе предварительной низкочастотной фильтрации.

Алгоритм скелетизации полутоновых изображений на основе адаптивной низкочастотной фильтрации

Скелеты часто используются в распознавании образов и, поэтому, должны быть стабильными при изменении контраста и действия шума. Эти свойства скелетов напрямую зависят от качества алгоритмов скелетизации. В условиях высокого контраста и слабого шума на исходном полутоновом изображении широко используются алгоритмы бинарной скелетизации.

Они относительно просты и могут быть устойчивы к мультипликативному шуму, проявляющемуся на границах областей после бинаризации. Однако, при снижении контраста и усилении зашумления исходного полутонового изображения скелеты, формируемые такими алгоритмами, разрушаются под действием аддитивного шума, проявляющегося в глубине областей скелетизируемого бинарного изображения.

Для скелетизации полутоновых изображений $\hat{I}$ с двухмодовой гистограммой яркости в условиях аддитивного шума предлагается алгоритм, основанный на адаптивных низкочастотной фильтрации и бинаризации для формирования скелетизированного изображения $S = \left\| \tilde{s}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ с использованием сглаженного $\tilde{I} = \left\| \tilde{i}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ и бинарного $B = \left\| \tilde{b}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ изображений . Сущность данного алгоритма состоит в выборе значения σ дисперсии Гаусс-фильтра, обеспечивающим минимальное значение модифицированной метрики чувствительности скелета к шуму, учитывающей искажения форм бинарных областей скелетизируемого изображения в результате низкочастотной фильтрации исходного полутонового изображения и вычисляемой с помощью выражения

$$M_{\mathrm{S}}(B,B,S) = \frac{1}{N_{\mathrm{S}}\left(S\right)} \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} m_{\mathrm{S}}(B,B,y,x), \tag{1}$$

где $N_{\mathrm{S}}\left(M_{\mathrm{B}}\right)$ – функция, определяющая количество единичных элементов в бинарной матрице $M_{\mathrm{B}} = \left\| m_{\mathrm{B}}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$, $N_{\mathrm{S}}\left(M_{\mathrm{B}}\right) = \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} m_{\mathrm{B}}(y,x)$; $m_{\mathrm{S}}(y,x)$ – чувствительность к локальным искажениям скелета, определяемая с помощью выражения

$$m_{\mathrm{S}}(B,B,y,x) = \begin{cases} 1 \text{ при } \left(N_{\mathrm{ES}}(y,x) > 2\right) \vee \left(\tilde{s}(y,x) \oplus \hat{s}(y,x) = 1\right), \\ 5 \text{ при } \left(\sum_{j=-1}^{1} \sum_{i=-1}^{1} \hat{b}(y+j,x+i) = 0 \right) \wedge \left(N_{\mathrm{ES}}(y,x) > 1\right), \\ 10, \text{ при } \left(\dfrac{\left| N_{\mathrm{S}}\left(B\right) - N_{\mathrm{S}}\left(B\right) \right|}{N_{\mathrm{S}}\left(B\right)} > 0{,}02 \right) \vee \left(R\left(S\right) \neq R\left(S\right) \right) \\ 0 \text{ в других случаях;} \end{cases} \tag{2}$$

где $R\left(B\right)$ и $R\left(B\right)$ – количество областей на изображениях B и B; $N_{\mathrm{ES}}(y,x) = 9 - \sum_{j=-1}^{1} \sum_{i=-1}^{1} \tilde{s}(y+j,x+i)$ – количество граничных пикселей для каждого пикселя скелета.

Алгоритм состоит из следующих шагов.

1) Инициализация значений переменных алгоритма: дисперсии ($\sigma = 1$); количества итераций ($n = 0$); матрицы бинарного изображения B $\left(B = f_{\mathrm{B}}\left(\hat{I}\right)\right)$.

2) Начало цикла вычисления метрики чувствительности. Формирование сглаженного изображения $\tilde{I}$ в результате свертки исходного изображения $\hat{I}$ с ядром $G(\sigma) = \left\| g(y,x,\sigma) \right\|_{\left(y=\overline{-Y_{\mathrm{G}},Y_{\mathrm{G}}},x=\overline{-X_{\mathrm{G}},X_{\mathrm{G}}},\right)}$ гаусс-фильтра, элементы которого определяются с помощью выражения

$$g(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{y^2+x^2}{2\sigma^2}}, \tag{3}$$

при $y = \overline{-\lceil 2\sigma \rceil, \lceil 2\sigma \rceil}$, $x = \overline{-\lceil 2\sigma \rceil, \lceil 2\sigma \rceil}$.

3) Формирование гистограммы яркости сглаженного изображения $\tilde{I}$ и определение по ней порога бинаризации T_{B}. Формирование бинарного изображения B с помощью алгоритма адаптивной пороговой обработки Otsu [4].

4) Формирование n-го скелетизированного изображения $S(n)$ с помощью алгоритма бинарной скелетизации (например ОРТА [5] и др.).

5) Вычисление n-го значения метрики $M_{\text{S}}(B,B,S,n)$ чувствительности скелета к шуму с использованием выражения (1).

6) Приращение значения дисперсии: $\sigma = \sigma + 1$.

7) Приращение счетчика количества итераций: $n = n + 1$.

8) Проверка счетчика количества итераций на достижение максимального значения, например, 20 (зависит от доступных вычислительных ресурсов). Если $n < 20$ переход на шаг 2.

9) Поиск минимального значения M_{SMin} метрики чувствительности скелета к шуму и номера N_{SMin} итерации для такой метрики: $M_{\text{SMin}} = \min\left(M_{\text{S}}(B,B,S,n)\right)$; $\left(M_{\text{SMin}} = M_{\text{S}}\left(B,B,S,n\right)\right) \Rightarrow \left(N_{\text{SMin}} = n\right)$ при $n = \overline{0,19}$.

10) Завершение алгоритма и формирование результата: $S\left(N_{\text{SMin}}\right)$.

Оценка эффективности алгоритмов скелетизации полутоновых изображений

Для оценки эффективности алгоритмов скелетизации использованы тестовые полутоновые изображения с контролируемым контрастом и уровнем аддитивного шума, формируемые согласно схеме, приведенной на рис. 1.

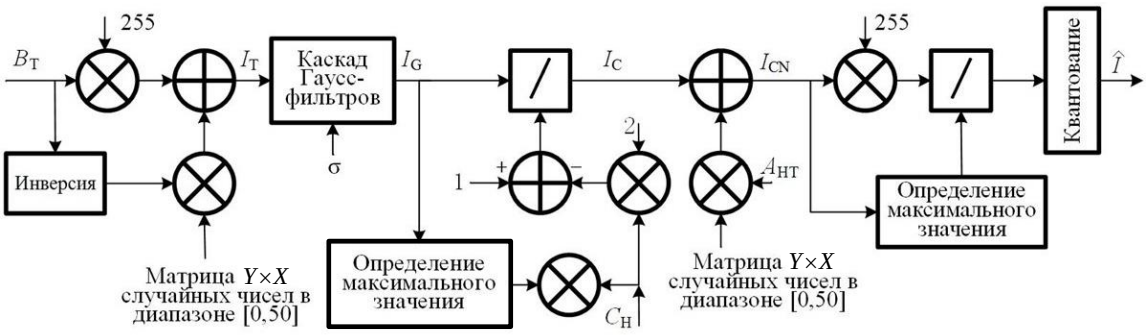


Рис. 1. Схема генерации тестовых полутоновых изображений с заданным контрастом и шумом с использованием бинарных изображений

На вход схемы подается тестовое бинарное изображение $B_{\text{T}} = \left\| b_{\text{T}}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$, на основе которого формируется тестовое полутоновое изображение $I_{\text{T}} = \left\| i_{\text{T}}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ с помощью выражения

$$i_{\text{T}}(y,x) = 255 b_{\text{T}}(y,x) + \text{Rand}(0,50)\overline{b_{\text{T}}(y,x)}, \qquad (4)$$

при $y = \overline{0,Y-1}$, $x = \overline{0,X-1}$,

где $\text{Rand}(n,m)$ – функция, формирующая случайное значение в заданном диапазоне $[n,m]$.

Изображение I_{T} проходит через три Гаусс-фильтра, в результате чего формируется размытое полутоновое изображение $I_{\text{G}} = \left\| i_{\text{G}}(y,x) \right\|_{\left(y=\overline{0,Y-1},x=\overline{0,X-1}\right)}$ с помощью выражения

$$i_{\mathrm{G}} = f_{\mathrm{G}}\left(f_{\mathrm{G}}\left(f_{\mathrm{G}}\left(i_{\mathrm{T}}(y,x),\sigma\right),\sigma\right),\sigma\right), \tag{5}$$

где $f_{\mathrm{G}}(I_{\mathrm{X}},\sigma)$ – функция свертки изображения I_{X} с ядром Гаусс-фильтра с параметром σ.

С помощью преобразования гистограммы яркости сглаженного изображения I_{G} с коэффициентом $C_{\mathrm{H}} \in (0;1)$ формируется низкоконтрастное изображение $I_{\mathrm{C}} = \left\| i_{\mathrm{C}}(y,x) \right\|_{(y=\overline{0,Y-1},x=\overline{0,X-1})}$, значения пикселей которого вычисляются с помощью выражения

$$i_{\mathrm{C}}(y,x) = \frac{i_{\mathrm{G}}(y,x)}{(1-2C_{\mathrm{H}})} + C_{\mathrm{H}} \max(I_{\mathrm{G}}), \tag{6}$$

при $y = \overline{0,Y-1}$, $x = \overline{0,X-1}$.

В изображение I_{C} добавляется аддитивный гауссовский шум, в результате чего формируется зашумленное низкоконтрастное полутоновое изображение $I_{\mathrm{CN}} = \left\| i_{\mathrm{CN}}(y,x) \right\|_{(y=\overline{0,Y-1},x=\overline{0,X-1})}$, значения пикселей которого вычисляются с помощью выражения

$$i_{\mathrm{CN}}(y,x) = i_{\mathrm{C}}(y,x) + A_{\mathrm{HT}}\mathrm{Rand}(-1,1), \tag{7}$$

при $y = \overline{0,Y-1}$, $x = \overline{0,X-1}$,

где A_{HT} – амплитуда шума.

На основе изображения I_{CN} формируется нормированное и квантованное тестовое изображение $\hat{I}$, значения пикселей которого вычисляются с помощью выражения

$$\hat{i}(y,x) = \left[\frac{255 i_{\mathrm{CN}}(y,x)}{\max(I_{\mathrm{CN}})} \right], \tag{8}$$

при $y = \overline{0,Y-1}$, $x = \overline{0,X-1}$,

где $[\ ]$ – символ операции округления до ближайшего целого.

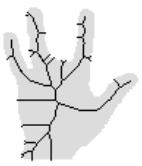
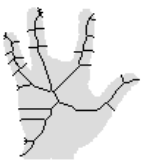
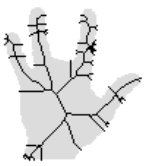
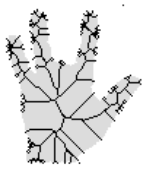
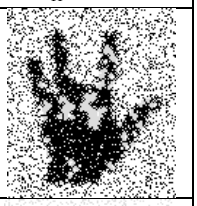
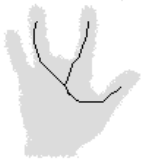
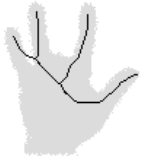
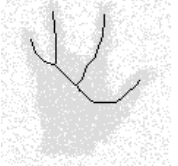
Алгоритм	$C_{\mathrm{H}} = 0,1$	$C_{\mathrm{H}} = 0,2$	$C_{\mathrm{H}} = 0,25$	$C_{\mathrm{H}} = 0,35$	$C_{\mathrm{H}} = 0,45$
OPTA					
ATF					
ATFM					

Рис. 2. Скелетизированные изображения, полученные при $\sigma = 1,0$, $A_{\mathrm{HT}} = 10$

На рис. 2 приведены комбинированные изображения ($B+\tilde{S}$), сформированные в результате сложения бинарных изображений B , полученных после адаптивной пороговой обработки, и бинарных скелетизированных изображений $\tilde{S}$, полученных с помощью алгоритмов OPTA, ATF и предложенного ATFM. Из рис. 1 следует, что алгоритм ATFM позволяет формировать скелеты более устойчивые к шуму (имеют меньше структурных искажений) по сравнению с OPTA и ATF.

Для полутоновых изображений с различными коэффициентом $C_{\text{Н}}$ модификации гистограммы, параметром размытием σ и амплитудой $A_{\text{НТ}}$ шума и различных алгоритмов в табл. 1 приведены значения отклонения E_{S} от эталонного скелета при отсутствии шума, вычисляемые с помощью разности изображений скелетов S и S_1 при отсутствии и наличии аддитивного шума ($S=S$ или $S=S$ в зависимости от алгоритма) в условиях заданного контраста:

$$E_{\text{S}}=\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left|S-\hat{S}\right|\Big/\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}S, \tag{9}$$

при $y=\overline{0,Y-1},\ x=\overline{0,X-1}.$

В табл. 1 приведены также значения контраста C_{I}, вариации V_{I}, адаптивного порога T_{B}, определяемого с помощью алгоритма Otsu, минимальное I_{MIN} и максимальное I_{MAX} значения пикселей полутонового изображения без шума и время T_{P} скелетизации. Значения контраста C_{I} и вариации V_{I} вычисляются с помощью выражения

$$C_{\text{I}}=\sqrt{\frac{1}{YX}\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left(\hat{i}\left(y,x\right)-\mu\left(\hat{I}\right)\right)^2}, \tag{10}$$

$$V_{\text{I}}=\frac{D_{\text{ST}}\left(I\right)}{\mu\left(I\right)}, \tag{11}$$

где $\mu\left(I\right)$ – среднее значение яркости пикселей для изображения I, $\mu\left(I\right)=\frac{1}{YX}\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}i\left(y,x\right)$;

$D_{\text{ST}}\left(I\right)$ – дисперсия значений пикселей для изображения I.

Табл. 1. **Оценка показателей эффективности алгоритмов скелетизации**

	$C_{\text{Н}}$	E_{S}			C_{I}	V_{I}	T_{B}	I_{MAX}	I_{MIN}	T_{P}, с		
		OPTA	ATF	ATFM						OPTA	ATF	ATFM
$A_{\text{НТ}}=10$ $\sigma=0,5$	0,1	0,005	0,631	0,151	0,326	0,799	125	230	30	0,225	3,089	2,855
	0,2	0,010	0,613	0,154	0,244	0,566	141	204	55	0,211	2,903	2,827
	0,25	0,106	0,695	0,318	0,203	0,459	151	191	68	0,212	2,962	2,870
	0,35	1,121	0,681	0,431	0,122	0,262	164	166	91	0,384	2,923	2,847
	0,45	11,773	0,766	0,631	0,040	0,082	180	140	115	1,066	4,048	4,058
$A_{\text{НТ}}=20$ $\sigma=0,5$	0,1	0,246	0,582	0,574	0,326	0,798	119	230	32	0,289	3,765	3,669
	0,2	0,968	0,713	0,441	0,244	0,567	119	204	55	0,443	3,848	3,741
	0,25	1,648	0,812	0,554	0,203	0,459	128	191	67	0,426	3,880	3,893
	0,35	5,249	0,649	0,585	0,122	0,263	134	166	91	0,972	5,119	5,061
	0,45	15,727	0,645	0,413	0,040	0,082	156	140	116	0,493	3,502	3,821
$A_{\text{НТ}}=10$ $\sigma=1$	0,1	1,074	0,635	0,356	0,313	0,766	128	230	38	0,373	3,861	3,767
	0,2	1,017	0,713	0,151	0,234	0,543	143	204	61	0,395	3,898	3,790
	0,25	1,342	0,663	0,156	0,195	0,441	146	191	72	0,389	3,866	3,785
	0,35	2,207	0,617	0,597	0,117	0,252	164	166	94	0,501	3,967	4,036
	0,45	12,663	0,713	0,587	0,039	0,079	182	140	116	0,995	4,497	4,812
$A_{\text{НТ}}=20$ $\sigma=1$	0,1	1,756	0,699	0,654	0,313	0,767	109	230	38	0,489	4,008	4,063
	0,2	2,305	0,064	0,515	0,234	0,543	115	204	58	0,538	3,962	4,142
	0,25	2,833	0,663	0,679	0,195	0,441	123	191	71	0,583	4,111	4,254
	0,35	6,049	0,667	0,649	0,117	0,251	141	166	94	0,971	5,893	6,347
	0,45	15,372	0,738	0,859	0,039	0,079	148	140	116	0,406	3,745	4,103

Для полутоновых изображений с резкими перепадами яркости (при $A_{\text{НТ}} = 20$) алгоритм ATFM позволяет: а) повысить устойчивость скелетов к шуму в сравнении с алгоритмом OPTA; б) при низком уровне шума ($A_{\text{НТ}} = 10$) и низком контрасте повысить устойчивость скелетов к шуму в сравнении с алгоритмом ATF; в) при высоком уровне шума ($A_{\text{НТ}} = 20$) сохранить устойчивость скелетов к шуму в сравнении с алгоритмом ATF. Предложенный алгоритм ATFM имеет одинаковую скорость скелетизации с алгоритмом ATF.

Для алгоритмов OPTA [5] и ZS [6] проанализированы зависимости отклонения E_{S} от отношения R_{SN} «сигнал/шум», вычисляемого с помощью выражения

$$R_{\text{SN}} = \frac{\frac{1}{YX}\sum_{y=0}^{Y-1}\sum_{x=0}^{X-1}\left(i(y,x) - I_{\text{MIN}}\right)}{D_{\text{ST}}\left(N_{\text{НТ}}\right)} \qquad (12)$$

При $\sigma = 0{,}5$, $0{,}17 > V_{\text{I}} > 0{,}16$ алгоритм ATFM позволяет уменьшить отклонения скелета от эталона при $R_{\text{SN}} = 2$ примерно в 3,6 раза в сравнении с алгоритмом OPTA (при проигрыше алгоритму ATF на основе OPTA в 1,2 раза) и в 2,1 раза в сравнении с алгоритмом ZS (при выигрыше по сравнению с алгоритмом ATF на основе ZS в 1,03 раза) и при $R_{\text{SN}} = 1$ примерно в 19,4 в сравнении с алгоритмом OPTA (при выигрыше по сравнению с алгоритмом ATF на основе OPTA в 1,2 раза) и в 12,5 раза в сравнении с алгоритмом ZS.

При $\sigma = 1{,}0$, $V_{\text{I}} \approx 0{,}65$, $R_{\text{SN}} = 5$ алгоритм ATFM позволяет уменьшить отклонение скелета от эталона примерно в 5,2, 2,8, 1,6 и 1,03 раз в сравнении с алгоритмами OPTA, ATF на основе OPTA, ZS и ATF на основе ZS соответственно. При $\sigma = 1{,}0$, $V_{\text{I}} = 0{,}164$ алгоритм ATFM позволяет уменьшить отклонение скелета от эталона при $R_{\text{SN}} = 2$ примерно в 4,8, 1,0, 2,7 и 1,1 раз в сравнении с алгоритмами OPTA, ATF на основе OPTA, ZS и ATF на основе ZS соответственно и при $R_{\text{SN}} = 1$ примерно в 16,7, 1,0, 11,3 и 1,06 раз в сравнении с алгоритмами OPTA, ATF на основе OPTA, ZS и ATF на основе ZS соответственно.

Заключение

Предложена модель скелетизации полутоновых изображений, основанная на двухмодовой гистограмме яркости и учитывающая влияние яркостно-контрастных параметров, мультипликативной и аддитивной составляющих шума на качество скелетизированного изображений. Модель позволяет определить условия эффективной скелетизации полутоновых изображений при которых искажения обусловлены влиянием только мультипликативной составляющей или комбинацией мультипликативной и аддитивной составляющих шума. Разработан алгоритм скелетизации полутоновых изображений основанный на предложенной модели и адаптивной низкочастотной фильтрации, отличающийся от известных алгоритмов учетом искажений форм областей скелетизируемого бинарного изображения в условиях аддитивного шума. Определены условия влияния контрастно-яркостных параметров изображений на качество скелетизации.

SKELETING OF LOW-CONTRAST NOISY HALFTONE IMAGES

J MA, A.A. BORISKEVICH

Abstract. To increase the stability of the skeletons of halftone images with a two-mode brightness histogram to noise, the article proposes a skeletonization model that considers the presence of multiplicative and additive noise components in a binary skeletonized image. Considering this model, a skeletonization algorithm has been developed, which considers the distortions in the shapes of the areas of the skeletonized binary image as a result low frequency filtering of the original halftone image and allows reducing errors in the skeletonization of halftone images.

Keywords: skeletonization of halftone images, multiplicative noise, additive noise, sensitivity of the skeleton to noise.

Список литературы

1. Hoffman M.E., Wong E.K. // Photonics West'98 Electronic Imaging, International Society for Optics and Photonics. 1998. Vol. 30. P. 1369–1373.
2. Chatbri H., Kameyama K. // Pattern Recognition. 2014. Vol. 42. P. 1–10.
3. Cai J. // The Computer Journal. 2012. Vol. 55. P. 887–896.
4. Otsu N. // IEEE trans. on systems, man, and cybernetics. 1979. Vol. 9. P. 62–66.
5. Chin R.T., [et al.]. // Computer Vision Graphics and Image Processing. 1987. Vol. 40. P. 30–40.
6. Zhang T.Y., Suen C.Y. // Communications of the ACM. 1984. Vol. 27. P. 236–239.

UDC C 620.9:.658.40

STRUCTURE AND COMPONENTS OF INTERNET OF THINGS NETWORK FOR IT PATIENT DIAGNOSTICS

U.A. VISHNYAKOU, H. TAO, Z. YIAN, W. HAORAN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 20, 2023

Abstract. Internet of Medical Things (IoMT) technologies increase the operational productivity and efficiency of healthcare organizations by optimizing clinical, information and operational processes. An overview of the use of IT development directions in remote diagnostics is given. For the development of IoT diagnostics (IoTD), an analysis was carried out in three directions: the iris, changes in voice characteristics, indicators of smart gadgets. The use of artificial intelligence technology and machine learning for IT diagnostics is considered. The structure of the IoTD patient diagnostic network is discussed.

Keywords: IoTD network concept, IT patient diagnostics iris, voice, smart gadget.

Introduction

The Internet of Medical Things (IoMT) is an Internet of Things network that includes medical devices (sensors), communicators, a server, software applications, and healthcare services. The interaction between sensors (devices) and the server allows healthcare organizations to optimize their clinical operations and workflow management, as well as improve patient care from remote locations [1, 2]. The main advantages of IoMT [1, 3]:

– rapid diagnosis: monitoring devices with intravenous infusion support constantly monitor the condition of patients, help diagnose diseases or health problems at an early stage;

– cost reduction: with the help of IoMT, the patient can receive high-quality medical consultations in real time without visiting doctors and hospitalization;

– improving access to health care in remote areas: villages and cities in some countries do not have sufficient access to modern medical services;

– reducing the number of manual errors: the data collected by connected medical devices does not allow errors, which improves the overall quality of diagnostics.

– improvement of medical care: since doctors can constantly monitor patients with medical devices based on intravenous infusion, there are more opportunities for specialized and individual treatment;

– effective management of medical equipment and medicines;

– detection of side effects of drugs in the early stages;

– medical insurance companies can use the data collected using connected medical devices to process claims, which ensures full transparency of their consideration.

The use of the Internet of Things (IoT) in healthcare makes it possible to move to a new level of disease diagnosis, treatment accuracy and monitoring of patients' health using micro- and nanodetectors and other «smart devices». Remote monitoring reduces the risks of unscheduled hospitalization and reduces the burden on hospitals, and the interaction between doctors and patients online is simplified.

Next, we will consider individual areas of medical diagnostics for building IoT diagnostics (IoTD).

IT gadget diagnostics

In 2008, a storm of «smart medical care» was set off around the world, and smart devices such as smart bracelets and smart robots began to emerge in public medical places such as hospitals [4]. Nowadays, medium-sized and large-scale hospitals and other public medical places have begun to upgrade and transform their medical systems intelligently, and apply advanced medical equipment [5] to the medical field of helping patients, which greatly improves the work efficiency of medical staff, saving a lot of manpower and material resources. Therefore, under the background that hospitals and other public medical places have been in the traditional working mode for a long time, and wearable devices are developing rapidly in the medical industry, this project researches and designs a patient data collection platform based on smart bracelets. The research value of this topic lies in that the smart bracelet can collect a variety of vital sign signals and apply them to patients in hospitals and other public health medical places. Its main significance and value are as follows:

1. The smart bracelet terminal hardware performs long-term, 2-hour uninterrupted real-time tracking and medical monitoring of vital sign signals of patients, and collects comprehensive and specific medical data of patients, so as to quickly find the cause, realize disease prevention and early medical treatment Favorable convenience is provided.

2. Integrate patients' data in the cloud, build medical servers and improve telemedicine-related hardware equipment, and carry out telemedicine treatment-related activities with tertiary hospitals and their medical experts and attending physicians. Realize the function from manual early warning to automatic early warning when the patient has an emergency. At this time, the system immediately feeds back the current status information of the patient to the medical experts and attending physicians, so as to realize the prevention of the disease and the high efficiency of medical treatment. The life, health and safety of seriously ill patients opens a green and intelligent acceleration channel to prolong the life cycle of patients.

3. When the patient is admitted to the hospital, the smart bracelet is distributed to the patient. When the patient is discharged, it is recycled, cleaned, disinfected and reused. It can be reused many times, reducing medical costs and avoiding the waste of medical resources. Therefore, it is very meaningful to research and design a patient data collection platform based on smart bracelets.

IT iris diagnostics

Modern iris medicine began at the end of the 18th century, and was first developed in Europe and North America. In modern iris medicine, the accuracy of the atlas is high, and it is widely used in the world. It has been recognized by iris doctors all over the world. The International Iridologist Practitioners Association, or IPA for short, is the largest international iris medicine organization in the world, which takes promoting and continuously improving the research and development of iris medicine as its mission. The development and application of iris medicine in developed countries have been widely recognized in recent years. Iris diagnostics is an effective means of preventive medicine. Through the detection of 800000 patients in Russia, it was found that the detection rate of iris medicine for diseases was as high as 85 % [6].

In recent years, iris medicine and iris diagnosis and treatment have been developing in various countries. Iris medicine (iris medicine, iridology, iris diagnostics, etc.), based on morphology, observes, predicts and infers the overall physique of the human body, the reality of the overall health, the occurrence and recovery of diseases by studying the morphological changes of the iris of the human eye, such as changes in color, color spots, structure, and pupil. It is a scientific and practical technology that reflects the genetic constitution of the human body and the defects of various organ systems, it is also a complete medical system including examination (detection), diagnosis (judgment and evaluation) and treatment (conditioning). The IoT for human health diagnosis based on iris data can provide more detailed and accurate diagnosis for iris medicine.

The purpose of first step is to detect the features corresponding to the established features in the reference image in the image to be registered. To this end, we will use various descriptors and similarity measures in the spatial domain composed of these features. Our second task is to find the correspondence of some feature points between two images. Since the transformation of iris texture is non rigid and

extremely irregular, which increases the difficulty of finding corresponding point pairs, here we use the method of Local jet model plus LBPs (local invariant binary patterns).

IT voice diagnostics

One application of IoT in healthcare is the use of speech analytics to diagnose human lung diseases. The Internet of Things (IoT) network for diagnosing human lung disease through speech analytics is an innovative approach to healthcare that uses IoT technology to collect and analyze patient data remotely. Once voice data is collected, machine learning algorithms can be used to analyze voice patterns and identify potential indicators of lung disease. For example, changes in pitch, tone and frequency can be used to detect the presence of lung diseases such as asthma, chronic obstructive pulmonary disease (COPD) and pneumonia. This approach works by using IoT-enabled devices, such as smartphones or wearables, where patients can record their voices and transmit the data to a central database for analysis. Machine learning algorithms can then be used to analyze voice patterns and identify potential indicators of lung disease. This approach enables remote monitoring of patients and eliminates the need for frequent in-person visits by healthcare providers. In addition to improving patient outcomes, an IoT network for diagnosing lung disease through speech analytics can reduce healthcare costs by minimizing the need for costly diagnostic tests and reducing hospital readmission rates.

However, there are some challenges with this approach. Including data accuracy, privacy and security issues, one issue is the potential for inaccurate or incomplete data. For example, if a patient does not speak clearly or has background noise, this may affect the accuracy of the diagnosis. Another issue is privacy and security. The collection of patient voice data raises concerns about data privacy and security. It is important to ensure the confidentiality and security of patient data and to take appropriate measures to prevent data leakage.

The proposed system is designed to make classifications and detect cough sounds [7]. There are four main stages after selecting the sound classification dataset. The first stage is extracting the features from audio files such as the MFFCs, chromagram, Mel-spectrogram, spectral contrast, and tonal centroid features. The second stage is labeling, it categorized the sound samples into cough and non-cough, then fed the inputs into a neural network. It reached the training stage and record the results until reached the optimal parameters according to the best results (changing epochs number, learning rate, etc.). The final stage, after generating the model, several tests were being applied on recorded sounds from volunteers.

Machine learning in IT patient diagnostics

The application of machine learning techniques in medical diagnostic analysis is becoming increasingly important. It works by building a model based on known information, and then using the model and related data to detect whether a patient has a specific condition or disease.

Machine learning techniques in diagnostic medical imaging hold great promise for detecting pathologies in images and identifying diseases present in images. It can effectively detect cancer brain damage, heart disease and other conditions, which require high-quality medical images to provide effective reference. In addition, it can detect earlier conditions, thereby helping to judge clinically earlier disease [8].

Functional imaging can help doctors accurately understand the pathological changes of patients and better judge the patient's condition. Machine learning technology can identify the characteristics of each patient and can provide effective tips for doctors, so that they can better judge the patient's disease.

The use of machine learning technology for medical diagnosis and analysis will bring more convenience to doctors, enabling faster and more accurate diagnosis of diseases, optimizing medical treatment experience, reducing possible diagnostic errors of doctors, thereby improving the efficiency and level of medical treatment, and better Ensure the health and safety of patients.

Structure of IoTD network

The structure of IoT diagnostics of patient (IoTD) network consists of the following components:

1. IoT devices are the physical devices that collect and transmit data. In the IoT diagnostics (IoTD) network, IoT devices can include smart phone, wearable devices, or other devices that can capture and transmit data.

2. Sensors are used to collect data about the patient's environment, such as temperature, humidity, and air quality. This information can be used to provide context for the data and improve the accuracy of the diagnosis.

3. Cloud or edge computing is used to process and analyze the data collected by the IoT devices. In the process of disease diagnosis, machine learning algorithms can be used to analyze disease characteristics and identify potential indicators of disease.

4. Data storage is used to store the disease data and other relevant information about the patient, such as medical history and demographic information. This information can be used to provide personalized care and improve the accuracy of the diagnosis.

5. User interface is used to provide healthcare providers with access to the patient data and diagnostic results. This can be a web-based interface or a mobile application.

6. Security and privacy measures are used to protect patient data and ensure that it is kept confidential and secure. This can include encryption, access controls, and other measures to prevent data breaches.

Overall, the Internet of Things Diagnostic (IoTD) network is structured to collect and analyze data from multiple sources to provide accurate diagnosis and personalized care to patients. In the IoT structure proposed by the author for the diagnosis of patients, data will be collected from patients' smartphones or from smart gadgets. These will be photos of the iris, the results of voice tests, parameters from gadgets (pulse rate, blood pressure, body temperature, etc.). The collected data will be transmitted to the IoT server for recording in the database, preliminary analysis, making decisions about the health of patients. The server will include deep neural networks trained on data sets of patients with certain diseases (by iris, voice markers, etc.). The results of the analysis will be sent to the smartphones of the attending doctors.

The application of machine learning techniques in medical diagnostic analysis is becoming increasingly important. It works by building a model based on known information, and then using the model and related data to detect whether a patient has a specific condition or disease.

Conclusion

The concept is given and the advantages of using IoT technologies in medicine are described, as well as the directions of IoT diagnostics development. To develop of IoT diagnostics approaches, an analysis of three directions was carried out: on the iris, indicators of smart gadgets, changes in voice characteristics. The use of artificial intelligence and machine learning technologies for the diagnosis of patients is discussed. The author's structure the IoT network for patient distance diagnostics is considered.

References

1. Aksenova E.I., Gorbatov S.Y. Internet of Medical Things (IoMT): new opportunities for healthcare. M.: GBU «NIIOZMM DZM», 2021.
2. Deloitte. Medtech and the Internet of Medical Things. How connected medical devices are transforming health care. [Electronic resource]. URL: https://www2.deloitte.com/global/en/pages/life-sciencesand-healthcare/articles/medtech-internet-of-medical-things.
3. Embitel. IoT in Healthcare – Connected Devices, Telemedicine and Remote Monitoring. [Electronic resource]. URL: https://www.embitel.com/blog/embedded-blog/iot-in-health-care-connected-devices-tele-medicine-and-remote-monitor-ing.
4. Fuqiang C. // Microprocessors and Microsystems. 2021. Vol. 82(5). P. 103901.
5. Shanguo L. // Microprocessors and Microsystems. 2021. Vol. 82(5). P. 103856.
6. Iridodiagnostics. [Electronic resource]. URL: https://mgkl.ru/patient/stati/iridodiagnostika.
7. Visniakou U.A., Shaya B.H. // Modern means of communication: materials of the 27th International Scientific Conference, Minsk, October 27−28. Minsk: BSAC. 2022. P. 29.
8. Kulkarni P. Machine learning in Medical Diagnoses. [Electronic resource]. URL: https://medium.com/ai-techsystems/ml-in-medical-diagnosis-1370b8ecfe31.

UDC [004.382.75-053.88]-027.31

DESIGN OF AUTOMATIC DISTRESS BRACELET FOR ELDERLY BASED ON SINGLE-CHIP MICROCOMPUTER

XU WEIXAUN, N.V. KHAJYNAVA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March13, 2023

Abstract. This article introduces the basic component of Automatic Distress Bracelet, as well as its functions, application scenarios. This device implements that when the user parameters are abnormal, the user address will be automatically sent to the user's home through SMS.

Keywords: microcomputer, IoT, automatic system.

Introduction

Every family has elderly people. Due to the decline of the body's physiological functions， elderly people prone to accidents and even life-threatening. It is conducive to better to build an elderly-friendly society that inform their families and ask passers-by for help in time, when the elderly is in danger.

This year, the concept of ensuring the safety of the elderly has become a social consensus. But products that can help to ensure the safety of the elderly have not gained popularity. Single-chip microcomputer can use to processing complex human body signals and has the advantages of low energy consumption and small size. Therefore, this is very suitable for achieving goals. What's more, different people have different physical indicators. It is the reason why it will always send the wrong distress signal if using the same value to define normal. Use the server and database to define the normal parameters of different users by process the parameters in the bracelet is one of the solutions.

This subject based in STC89C52 as processor and supplemented by several modules for example SKB360, GSM and Thermometer. When user's parameters are abnormal, it will get location information via satellite and send the location to the user's family.

Positioning system section

In this section, it is necessary to obtain the location information and send the obtained data to the specified mobile phone number.

The main parts of this system are shown in Figure 1.

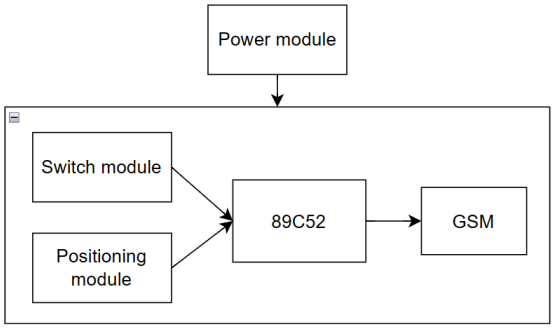


Figure 1. Block diagram of positioning system

The reason of why STC89C52 microcontroller is used, as the core is STC89S52 is a low-power, high-performance CMOS8-bit microcontroller with 8K in-system programmable Flash memory.

The microcontroller has low power consumption, rich interfaces, and low cost, which can fully meet the requirements of this design.

First of all, the normal operation of the STC89C52 is inseparable from the SCM minimum system, which consists of an STC89C52 and three parts: reset circuit, clocking circuit and power supply circuit. We can easily find the relevant circuit online. Figure 2 shows one of these.

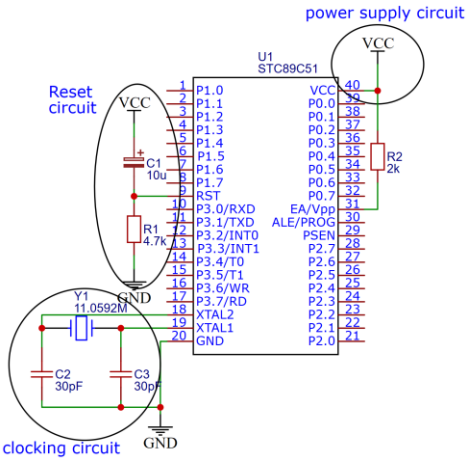


Figure 2. Example of a SCM minimum system

Then, a positioning device should be used to determine the location. NEO-6M GPS is a very popular, cost-effective and high-performance GPS module with a ceramic SMD antenna, an on-board memory chip, and a backup battery, which can be easily integrated with various microcontrollers [1].

When using this module, we do not need to have a clear understanding of the internals. The diagram of the equivalent interface of NEO-6M GPS is shown in Figure 3.

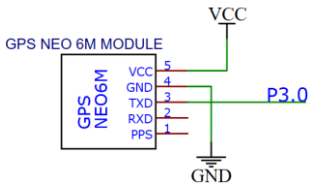


Figure 3. Diagram of the equivalent interface

After the navigation system is encapsulated as shown in the Figure 3, only consider five pins. By connecting TXD to the RXD of the microcontroller, positioning information can be obtained by using command like $GPRMC.

The sending function of information is implemented by the GSM module. The GSM module has all the basic functions for communication based on GSM network. In other words, simply put the GSM module plus the keyboard, display and battery is a mobile phone.

SIM800A is a GSM/GPRS module from SIMCOM, which size is 24×24×3mm and can be applied to various compact product design needs. The module's high sensitivity, low power consumption and lightweight size make it suitable for automotive, handheld devices such as PDAs, vehicle surveillance, mobile phones, cameras and other mobile positioning system applications. Therefore, it is fully meet the requirements of this design [2–4].

Similarly, the SIM800A module has is encapsulated by using SMT. It can simply be thought of as a module with only four pins. The diagram of the equivalent interface of SIM800A is shown in Figure 4.

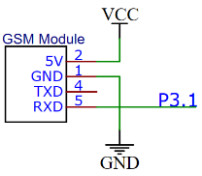


Figure 4. Diagram of the GSM Module

Of course, after successfully connecting the SIM800A module and the microcontroller, a SIM card is also required to use the network provided by the operator to operate.

Switching system section

This section is used to decide whether to send user's address via SMS. In other words, it is the Switching module of Positioning system. Here will use a thermometer as an example to illustrate the composition of this system.

First of all, it is necessary to discuss the usefulness of uploading data to the cloud. The data is saved to better monitor the state of the body, which is very helpful in predicting and preventing certain diseases. Moreover, the system should have a certain learning ability that adjust the standards according to the user's situation.

The size of the bracelet determines that it cannot store a lot of data. And, the bracelet is not a good user-friendly interface. Therefore, it is necessary to use the cloud. The main parts of Switching system are shown in Figure 5.

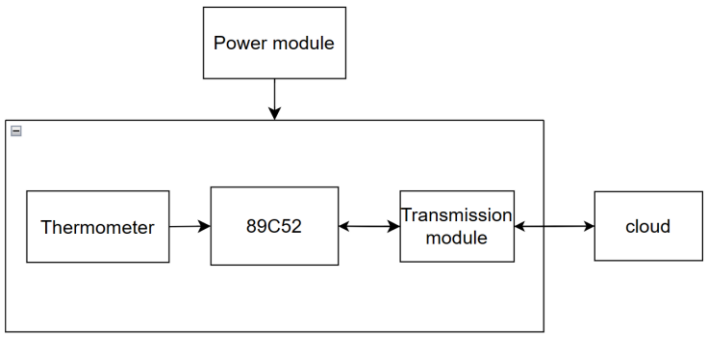


Figure 5. Block diagram of switching system

There are two main ways to transmit and save the data collected by smart bracelets.

The first is to connect the bracelet to the network by WIFI. The ESP8266 module can be used to achieve this effect. This method requires additional steps to identify the user of these bracelets. However, it no additional action is required after successful identification.

The second is to connect with bracelet and phone by Bluetooth. This method requires Occasional manual manipulation. However, this design is simpler and the supporting mobile phone software can provide users with a more beautiful interface. It is a convenient method to logotype user that user only need to log in to the client and connect the bracelet via Bluetooth on his phone.

Finally, learning is the process of grasping patterns in data and improving oneself or making predictions. By identifying a large number of parameters and summarizing the characteristics through algorithms, the system can achieve the purpose of parameter adjustment. [5] Of course, there may be an alarm when the parameters are normal in the early stage. Therefore, there must be a design to cancel the alarm.

Conclusion

The complete bracelet system consists of a hardware system and a software system. This report focuses on the design of hardware systems. Because the whole program is more complex, and the amount of calculation is large, and more floating-point number calculations are used, it is recommended to use C language for the writing of the program for easy reading.

References

1. Guide to NEO-6M GPS Module with Arduino [Electronic resource]. URL: https://randomnerdtutorials.com/guide-to-neo-6m-gps-module-with-arduino/
2. GPS Module [Electronic resource]. URL: https://www.datasheetq.com/CSR8635-doc-CSR/
3. GPS Module [Electronic resource]. URL: https://blog.csdn.net/R_Z_Q/article/details/104464836/
4. SIM800A [Electronic resource]. URL: https://cn.simcom.com/product/SIM800A.html
5. Haykin // Neural Networks: A comprehensive Foundation, 2nd Edition. Vol. 2. P. 43–44.

УДК 621.391

ПРОГРАММНАЯ МОДЕЛЬ ДВИЖЕНИЯ УЗЛОВ САМООРГАНИЗУЮЩЕЙСЯ СЕТИ В ТРЕХМЕРНОМ ПРОСТРАНСТВЕ

Т.В. ПОЛУЯН, С.Н. КАСАНИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Беспроводные мобильные самоорганизующиеся сети (МСОС) состоят из динамических узлов, имеющих высокую скорость движения. По сравнения с другими существующими видами сетей связи, в МСОС возникает больше всего проблем при организации маршрутизации, так как ввиду высокой скорости движения узлов значительно усложняется задача качественной доставки передаваемых пакетных данных. Под качеством подразумевается целостность доставленных пакетов и минимизация задержек. В статье описывается разработанная программная модель движения узлов, учитывающая оценку времени обслуживания каналов по расстоянию, относительную скорость и ускорение между узлами, необходимые для оптимизации построения транспортных таблиц и повышения качества связи.

Ключевые слова: мобильные самоорганизующиеся сети, маршрутизация, программная модель, скорость, качество связи, пакетные данные.

Введение

В связи с тенденцией к развитию технологий непилотируемых летательных аппаратов увеличивается количество беспилотной техники как гражданского, так и специального назначения. Обычно наземные станции, летающие узлы и спутники состоят из трехслойной сети. Летающие узлы сети занимают средний слой такой сети, связь с глобальной паутиной осуществляется через каналы связи с наземными станциями. Однако, в случае выхода узла из зоны действия наземной станции, передача пакетных данных будет осуществляться через спутник, что требует больших затрат и высокой задержки (около 250 мс). Альтернативой данному способу связи является использование в качестве отправителя и получателя сообщений самих сетевых узлов, таким образом, узлы будут являться и ретрансляторами для формирования самоорганизующейся сети пересылки данных с несколькими переходами. Таким образом, анализ маршрутизации в мобильных самоорганизующихся сетях (МСОС) является актуальной и необходимой тематикой современных исследований.

Многие исследования были посвящены протоколам и алгоритмам маршрутизации высокодинамичных МСОС. В [1] представлен метод оценки надежности канала, учитывающий возможное время обслуживания канала и непредсказуемость топологических изменений МСОС. Реактивный протокол маршрутизации на основе местоположения MUDOR, предложенный в [2], использует доплеровский сдвиг, который представляет собой относительную скорость, указывающую, насколько быстро узлы перемещаются близко или далеко друг от друга, для оценки стабильности канала. В [3] моделируется система с географической маршрутизацией в рамках нескольких моделей мобильности и представлены две схемы прогнозирования мобильности (ПМ), учитывающие анализ влияния ошибок определения местоположения, вызванных движением узла. Еще один из рассматриваемых подходов [4] описывает схемы мобильности с учетом потерь многоузловых ретрансляторов (МУР) и маршрутизации в случае высокой динамики узлов. При прогнозировании выхода за пределы диапазона передачи МУР или следующего перехода узел пересчитывает набор МУР или таблицу маршрутов.

Все упомянутые выше протоколы маршрутизации улучшают производительность МСОС, выбирая в качестве маршрутов более стабильные каналы связи. Однако основное предположение этих улучшений заключается в том, что скорость узлов постоянна, хотя она быстро меняется в высокодинамичных сценариях.

Целью работы является определение требований к структуре и основных элементов программной модели движения узлов мобильной самоорганизующей сети.

Сетевая модель

Сетевая модель является основой для разработки программной модели движения узлов. Рассматривается физический уровень сетевой модели, на котором работают мобильные узлы. Сеть содержит *n* узлов, обозначенные как множество *N*. Два узла сети обозначены как *a* и *b*. Когда два узла находятся в условиях радиовидимости, существуют линия связи между этими узлами l_{ab}. Набор доступных линий связи в сети обозначает как *L*. Если существует линия l_{ab}, принадлежащая к *L*, узел *b* является соседним с узлом *a*, множество обозначается как N_a^1. Так же, если существует линия l_{ad}, принадлежащая к *L*, в то время как l_{ad} не принадлежит к N_a^1, можно сказать, что *d* является одним из двух соседей узла *a*, который обозначен как N_a^2.

Дальнейшая разработка проводится, основываясь на следующих предположениях:

1. Каждый узел имеет одинаковую мощность передачи и радиус действия.
2. Все каналы связи в модели сети однородны.
3. Все узлы в сети следуют одной и той же модели движения.
4. Узлы движутся независимо друг от друга.

Модель движения

На точность прогнозирования времени обслуживания канала влияет модель подвижности узлов, в то время как движение узлов предполагается равномерно-переменным. Диапазон скорости и координат узлов ограничен в определенной области, начальные координаты расположения узлов заданы, а скорость и направление движения формируются случайным образом во время начала движения для всех узлов. Узлы движутся с учетом следующих условий:

1. Движение каждого узла является равномерно-переменным.
2. Скорость изменяется в диапазоне $[\upsilon_{min}, \upsilon_{max}]$.
3. Направления движения находится в диапазоне $[0, 2\pi)$.
4. Направление вектора ускорения совпадает с направлением вектора скорости, когда скорость достигает предельного значения, ускорение принимает нулевое значение.
5. На каждом временном интервале узлы определяют ускорение текущего движения. Вероятность равномерного, ускоренного или замедленного движения узлов составляет p_1, p_2, p_3, причем, $p_1, p_2, p_3 \in [0,1]$ и $p_1 + p_2 + p_3 = 1$.
6. Перемещение узлов ограничено трехмерной областью $X \times Y \times Z$, за исключением отрицательных значений координатной оси Z, обозначающей высоту пространства.

Модель взаимодействия узлов

Расстояния между узлами сети рассчитывается аналогично, как и между двумя описанными выше узлами *a* и *b*, модель взаимодействия которых представлена на рис. 1. Расстояние между узлами обозначается как $a\,r_a$ и r_b, вектор расстояния – $\overrightarrow{r_{ab}}$. Таким образом, модуль $\overrightarrow{r_{ab}}$ вычисляется как

$$\left|\overrightarrow{r_{ab}}\right| = \sqrt{\left(x_a - x_b\right)^2 + \left(y_a - y_b\right)^2 + \left(z_a - z_b\right)^2}\,, \qquad (1)$$

где $x_a, x_b, y_a, y_b, z_a, z_b$ являются координатами узлов.

Вектора скорости и ускорения двух узлов обозначены как $\overrightarrow{\upsilon_a}, \overrightarrow{a_a}, \overrightarrow{\upsilon_b}, \overrightarrow{a_b}$. Аналогично вектора относительной скорости и ускорения между узлами обозначается как $\overrightarrow{\upsilon_{ab}}, \overrightarrow{a_{ab}}$. Модуль $\left|\overrightarrow{\upsilon_{ab}}\right|$ вычисляется как

$$\left|\overrightarrow{\upsilon_{ab}}\right| = \sqrt{\left(\upsilon_{bx} - \upsilon_{ax}\right)^2 + \left(\upsilon_{by} - \upsilon_{ay}\right)^2 + \left(\upsilon_{bz} - \upsilon_{az}\right)^2}, \quad (2)$$

где $\upsilon_{bx}, \upsilon_{ax}, \upsilon_{by}, \upsilon_{ay}, \upsilon_{bz}, \upsilon_{az}$ – проекции векторов на оси X, Y, Z.

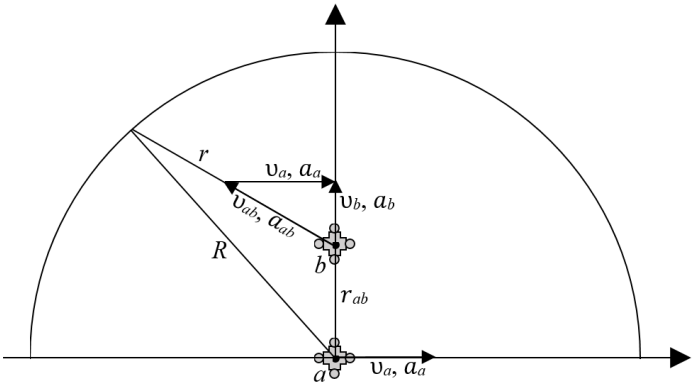


Рис. 1. Модель взаимодействия узлов

Программная модель движения

Программная модель разработана с учетом вышеизложенных положений сетевой модели, моделей движения и взаимодействия узлов. Алгоритм работы программной модели следующий.

1. Начальное положение сетевых узлов задано по умолчанию в одной точке, дальнейшее расположение выбирается рандомно в пределах заданной трехмерной плоскости относительно исходной точки.

2. По полученным координатам рассчитываются расстояния между узлами сети. Генерируется матрица расстояний в заданные моменты времени.

На рис. 2 приведена UML-диаграмма классов программной модели, показывающая соответствие принципам объектно-ориентированного программирования и требованиям к разработке программного обеспечения.

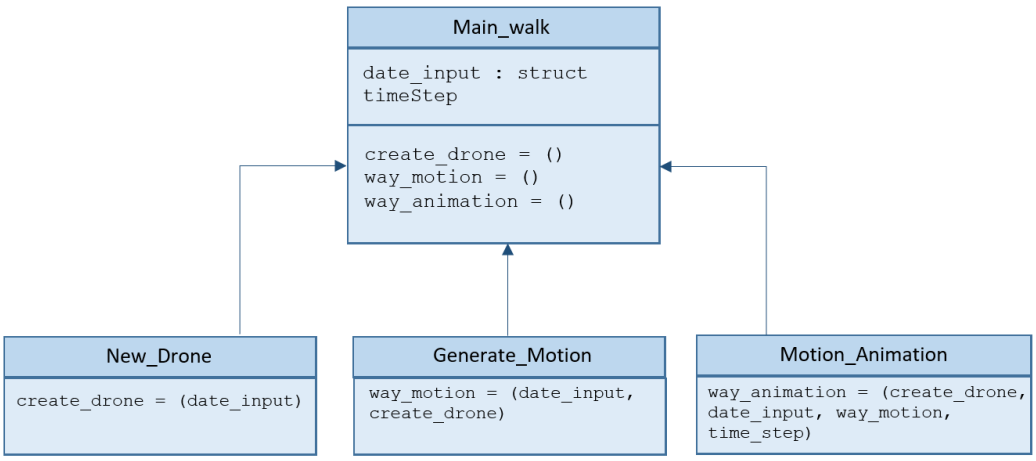
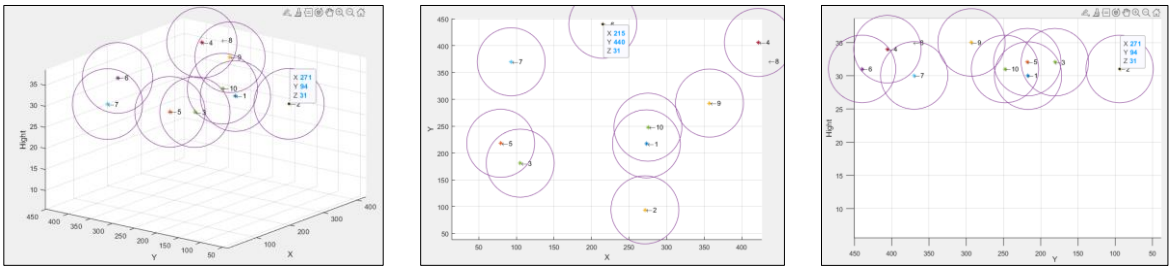


Рис. 2. UML-диаграмма классов программной модели

Трехмерная модель построена таким образом, чтобы при разработке алгоритма маршрутизации и построения маршрутных таблиц можно было учесть все параметры изменения расположения узлов, такие как координаты узла в момент времени (в том числе высоту),

расстояние между ближайшми соседями и всеми узлами в канале передачи, рассчитать скорость и направление движения узлов и ускорение. Перечисленные базовые параметры необходимы для прогнозирования изменения сетевой топологии, что немаловажно для построения оптимального маршрута передачи данных.

На рис. 3 приведы результаты моделирования с учетом радиовидимости сетевых узлов.



а *б* *в*

Рис. 3. Пример расположения сетевых узлов в трехмерном пространстве:
а – 3D плоскость; *б* – вид сверху; *в* – вид сбоку

Заключение

Разработана программная модель движения сетевых узлов, включающая в себя сетевую модель, модель движения и модель взаимодействия узлов. Получены матрицы расстояний между узлами с заданным интервалом времени с учетом относительной скорости и ускорения движения узлов. Благодаря гибкости и изменяемости, разработанная программная модель может применяться для оценки качества связи в высокодинамичных мобильных сетях, разработки методов увеличения скорости передачи пакетов данных и минимизации потерь информации.

MOTION SOFTWARE MODEL OF SELF-ORGANIZING NETWORK NODES IN THREE-DIMENSIONAL SPACE

T.V. POLUYAN, S.N. KASANIN

Abstract. Wireless mobile self-organizing networks (MSOS) consist of dynamic nodes with a high speed of movement. Compared with other existing types of communication networks, in MSOS the most problems arise when organizing routing, since due to the high speed of movement of nodes, the task of high-quality delivery of transmitted packet data becomes much more complicated. Quality refers to the integrity of delivered packets and the minimization of delays. The article describes the developed software model of the nodes movement, which takes into account the estimation of the service time of channels by distance, the relative speed and acceleration between nodes, necessary to optimize the construction of transport tables and improve the quality of communication.

Keywords: mobile self-organizing networks, routing, software model, speed, communication quality, packet data.

Список литературы

1. Lei Lei, Wang Dan // Link availability estimation based reliable routing for aeronautical ad hoc networks, Ad Hoc Networks. 2014. Vol. 20. P. 53–63.
2. Biomo J.M.M., Kunzand St-Hilaire M.T. // 2014 7th IFIP Wireless and Mobile Networking Conference (WMNC). 2014. P. 1–7.
3. Son D., Helmy A., Krishnamachari B. // IEEE Transactions on Mobile Computing. 2004. P. 233–245.
4. Sharma S. // IEEE 34th Conference on Local Computer Networks. 2009. P. 237–240.

UDC 004.421

HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA AND CONVOLUTIONAL NEURAL NETWORK

Z. WAN, A.A. BARYSKIEVIC

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 18, 2023

Abstract. With a widespread of various sensors embedded in mobile devices, the analysis of human daily activities becomes more common and straightforward. Human activity recognition (HAR) is a prominent application of advanced Machine Learning (ML) and Artificial Intelligence (AI) techniques that utilizes computer vision to understand the semantic meanings of heterogeneous human actions. This paper describes a supervised learning method that can distinguish human actions based on data collected from practical human movements. The primary challenge while working with HAR is to overcome the difficulties that come with the cyclostationary nature of the activity signals. This study proposes a HAR classification model based on a Convolutional Neural Network (CNN) and uses the collected human action signals. The model was tested on the WISDM dataset, which resulted in a 92 % classification accuracy. This approach will help to conduct further researches on the recognition of human activities based on their biomedical signals.

Keywords: human activity recognition, machine learning, convolutional neural network.

Introduction

Humans possess an amazing skill to comprehend information that others pass on through their movements like the gesture of a certain body part or the motion of the entire body. We can differentiate among human postures, track complex human motions, and evaluate human-object interactions to realize what they are doing, and even deduce what they intend to do. Even though these are advanced recognition functionalities performed by the brain based on the images of the surroundings captured by the eyes, the process occurs almost autonomously to us. Machines, on the other hand, are still learning how to apprehend various human activities, and we are teaching them based on our knowledge and understandings of the task. Considering the fact that machines (or computers) were nothing but simple calculators to solve arithmetic problems just sixty years ago, their understanding of complex concepts has come a long way. ML as a part of the AI, has given machines the capacity to interpret various situations in their surroundings and respond accordingly like humans. HAR is being researched since the early 1980s because of its promise in many applied areas. However, the significant breakthroughs in this field have come within the last two decades [1]. The recent developments in microelectronics, sensor technology, and computer systems have made it possible to collect information that is more fundamental from human movements, and the advanced ML techniques have made that information more comprehensible to the machines.

There are several approaches to collect HAR data from the participating subjects; broadly, they fall into one of the two categories – namely camera-based recording or sensor-based recording [2]. In the former approach, one or more video cameras are set up to record the activities of a subject for a certain amount of time, and then the recognition is performed using video analysis and processing techniques. The later one utilizes various types of sensors to track the movements of the subject. This approach can be further classified based on the type of sensors used, whether they involve wearable body sensors or the external ones [1]. External sensors are placed in predetermined points of interest on the subjects' body, whereas wearable sensors require to be attached to the subject while collecting data. Each of these techniques has its advantages, shortcomings, and apposite applications. Some recognition techniques even combine multiple recording techniques to collect data that are more relevant and make the corresponding actions more interpretable to the machines. The applications of HAR

include intelligent surveillance, haptics, human-computer interaction, motion or gesture-controlled devices, automatic health-care monitoring systems, prosthetics, and robotics. Despite many advancements, HAR is still a challenging task because of the articulated nature of human activities, the involvement of external objects in human interactions, and complicated spatiotemporal structures of the action signals [3]. Success in recognizing these activities requires advanced signal and image processing techniques, as well as sophisticated ML algorithms. Since the absolute performance is yet to be achieved, HAR remains a tending field to the researchers.

Datasets

The WISDM dataset contains mobility information that was collected from 30 people of different ages (ranging from 19 to 48 years), genders, heights and weights using a wrist-mounted smartphone. The smartphone has integrated accelerometer and gyroscope. Action data was recorded using these sensors while each of the subjects was performing six predefined tasks, which according to the jargon of ML, represent six different classes. Three-axial linear acceleration and three-axial angular velocity data were acquired at a steady rate of 20 Hz. The collected samples were labeled manually afterward. Before putting in the dataset, the samples were pre-processed using a median filter for noise cancellation and a thirdorder low-pass Butterworth filter having a 20 Hz cutoff frequency.

The proposed physical activity recognition algorithm

This study aims to classify the HAR signals of the WISDM dataset employing a CNN model, as shown in Figure 1. The training stage requires a set of data samples containing various attributes measured from subjects while performing various predefined activities. The supervised learning technique then try to make some "sense" out of the data, find out how the samples that belong to the same class are similar to each other while samples from different classes are diverse, then builds one or more internal models focusing on the crucial attributes that can highlight those contrasting properties to carry out the classification [1]. In the training stage, a preordained portion of the dataset is used to train the machine and build a feasible model, which is then evaluated over the remaining samples.

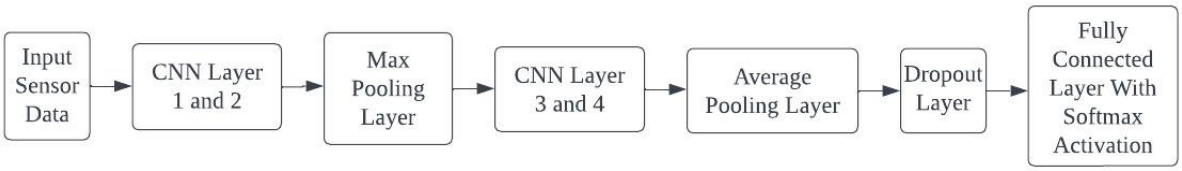


Figure 1. Block diagram of the proposed CNN-based HAR algorithm

The data has been preprocessed in such a way that each data record contains 80 time slices (data was recorded at 20 Hz sampling rate, therefore each time interval covers four seconds of accelerometer reading). Within each time interval, the three accelerometer values for the x axis, y axis and z axis are stored. This results in an 80×3 matrix. The data must be passed into the neural network as a flat vector of length 240. The first layer in the network must reshape it to the original shape, which was 80×3. The model is tested on the human behavior pose dataset WISDM. We selected 80 % of the data in the WISDM dataset for model training and 20 % for model testing. Using Adam as optimizer, batch size equals to 400, epochs equal to 50 (training for 50 rounds).

CNN Layer 1

The first layer defines 100 filters of height 10. This allows us to train 100 different features on the first layer of the network. The output of the first neural network layer is a 71×100 neuron matrix. Each column of the output matrix holds the weights of one single filter. With the defined kernel size and considering the length of the input matrix, each filter will contain 71 weights. The structure and parameters are shown in Table 1.

Table 1. **The input, condition and output of CNN layer 1**

Modules	Value
Input	Two order tensor (1086393,6)
	dtype: float32
Condition	Weight matrix size [10,100]
	Bias vector [10]
	activation function: Relu
Output	Two order tensor (100,71)
	dtype: float32

CNN Layer 2

The result from the first CNN layer will be fed into the second CNN layer. We will again define 100 different filters to be trained on this level. Following the same logic as the first layer, the output matrix will be of size 62×100. The structure and parameters are shown in Table 2.

Table 2. **The input, condition and output of CNN layer 2**

Modules	Value
Input	Two order tensor (100,71)
	dtype: float32
Condition	Weight matrix size [10,100]
	Bias vector [10]
	activation function: Relu
Output	Two order tensor (100,62)
	dtype: float32

Max pooling layer

A pooling layer is often used after a CNN layer in order to reduce the complexity of the output and prevent overfitting of the data. We chose a size of 3, which means that the size of the output matrix of this layer is only a third of the input matrix. The structure and parameters are shown in Table 3.

Table 3. **The input, condition and output of max pooling layer**

Modules	Value
Input	Two order tensor (100,62)
	dtype: float32
Condition	Weight matrix size [3,64]
	Bias vector [64]
Output	Two order tensor (100,20)
	dtype: float32

CNN Layer 3 and 4

Another sequence of 1D CNN layers follows in order to learn higher level features. The output matrix after those two layers is a 2×160 matrix. The structure and parameters are shown in Table 4.

Table 4. **The input, condition and output of CNN layer 3, 4**

Modules	Value
Input	Two order tensor (100,20)
	dtype: float32
Condition	Weight matrix size [10,100]
	Bias vector [10]
	activation function: Relu
Output	Two order tensor (160,2)
	dtype: float32

Average pooling layer

One more pooling layer to further avoid overfitting. This time not the maximum value is taken but instead the average value of two weights within the neural network. The output matrix has a size of 1×160 neurons. Per feature detector there is only one weight remaining in the neural network on this layer. The structure and parameters are shown in Table 5.

Table 5. **The input, condition and output of average pooling layer**

Modules	Value
Input	Two order tensor (160,2)
	dtype: float32
Condition	Weight matrix size [2,1]
Output	Two order tensor (160,1)
	dtype: float32

Dropout layer

The dropout layer will randomly assign 0 weights to the neurons in the network. Since we chose a rate of 0,5, 50 % of the neurons would receive a 0 weight. With this operation, the network becomes less sensitive to react to smaller variations in the data. It should further increase our accuracy on unseen data. The output of this layer is still a 1×160 matrix of neurons. The structure and parameters are shown in Table 6.

Table 6. **The input, condition and output of dropout layer**

Modules	Value
Input	Two order tensor (160,1)
	dtype: float32
Condition	Rate = 0.5
	Bias vector [64]
Output	Two order tensor (160,1)
	dtype: float32

Fully connected layer with Softmax activation

The final layer will reduce the vector of height 160 to a vector of 6 since we have six classes that we want to predict (Jogging, Sitting, Walking, Standing, Upstairs, Downstairs). This reduction is done by another matrix multiplication. Softmax is used as the activation function. It forces all six outputs of the neural network to sum up to one. The output value will therefore represent the probability for each of the six classes. The structure and parameters are shown in Table 7.

Table 7. **The input, condition and output of fully connected layer**

Modules	Value
Input	Two order tensor (160,1)
	dtype: float32
Condition	Weight matrix size [160,6]
	Bias vector [6]
Output	Two order tensor (6,1)
	dtype: float32

Results

Our approach to classifying the samples of six different classes contained in it, as well as the techniques and methods that we have employed in the proposed methodology. We set a classification model where the provided training samples were used to train the two-channel CNN model, and the rest of the samples were used to test it. The result yields a classification accuracy of 92 % on the test samples. Figure 2. presents the classification accuracies on both the training and testing samples at each epoch.

As seen in the figure, the training accuracy gradually increased with each epoch. The performance of the model was slightly unstable throughout the first 20 epochs, but it became pretty stable afterward.

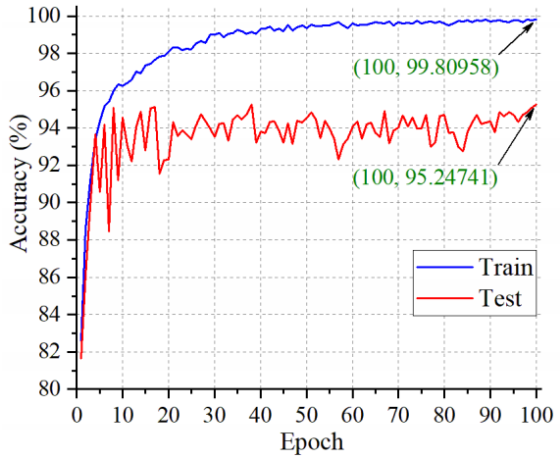


Figure 2. Train and test classification accuracies at each epoch

The confusion matrix provides more details on the output of the classification process. Figure 3. provides the confusion matrix of the epoch of our model for HAR classification. It is apparent that the model works very well while distinguishing six classes (Walking, Upstairs, Downstairs, Sitting, Standing and Jogging) registering over 95 % individual classification accuracies for each class.

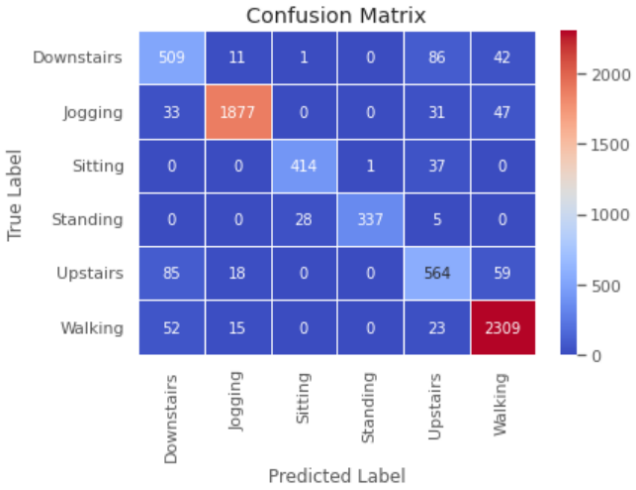


Figure 3. Confusion matrix of the HAR classification

Accuracy: For a given test dataset, the ratio of the number of samples correctly classified by the classifier to the total number of samples is the correct rate for the identified samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

where TP means True Positives, FP means False Positives, FN means False Negatives, and TN means True Negatives.

The model learns well with accuracy reaching above 92 % and loss hovering at around 0,39.

Conclusion

A CNN-based HAR classification model is proposed in the paper. It is tested on the WISDM dataset. The obtained results yield a 92 % classification accuracy. However, the model can be further modified by tuning specific parameters of CNN and adding more nodes and layers in the CNN architecture. A new set of features can also be extracted and fed in an additional channel of CNN to improve the model’ s performance, which is subjected to future studies.

References

1. Labrador M., Lara Yejas O. // Human activity recognition: using wearable sensors and smartphones. CRC Press, 2013.
2. Fu Y. // Human activity recognition and prediction. Springer, 2016.
3. Wang J., Liu Z. // Human Action Recognition with Depth Cameras. Cham: Springer International Publishing, 2014.

УДК 621.391

УСИЛЕНИЕ ОПТИЧЕСКОЙ НЕСУЩЕЙ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРИЕМА В ВОЛОКОННО-ОПИЧЕСКИХ СИСТЕМАХ ПЕРЕДАЧИ

Я.В. РОЩУПКИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. В работе произведено моделирование волоконно-оптической системы передачи с повышенной эффективностью приемного модуля за счет избирательного усиления оптического несущего колебания и трансформации спектра передаваемого сигнала. Обоснована и оценена структура приемного оптического модуля. Определены оптимальные параметры волоконно-оптического усилителя и схем последетекторной обработки сигнала.

Ключевые слова: волоконно-оптическая система передачи, приемный оптический модуль, вынужденное рассеивание Мандельштама-Бриллюэна, оптический усилитель.

Введение

Для повышения отношения сигнал-шум (ОСШ) и как следствие дальности связи в волоконно-оптических системах передачи можно использовать метод избирательного усиления оптической несущей с предварительной трансформацией спектра передаваемого сигнала, который подробно рассматривается в работе [1]. Основой данного метода является усиление несущего колебания принимаемого оптического сигнала при помощи оптического усилителя, который должен обеспечивать усиление только узкой полосы спектра в окрестности несущей. Большинство существующих волоконно-оптических усилителей, в том числе и широко распространенные EDFA, имеют очень широкую полосу усиления, достигающую нескольких терагерц, что неприемлемо для избирательного усиления. Наилучшим образом для данных целей подходит распределенный оптический усилитель на эффекте вынужденного рассеивания Мандельштама-Бриллюэна, который обладает узкой полосой усиления (десятки МГц), значительным коэффициентом усиления (более 30 дБ) и где в качестве усилительной среды выступает сама оптическая линия связи [2].

На передающей стороне требуется осуществить перенос спектра информационного сигнала из области низких частот в область более высоких, с тем чтобы в процессе оптической модуляции создать некоторый защитный интервал между несущей и спектральными составляющими информационного сигнала [3]. Этого можно достичь предварительным переносом информационного сигнала на поднесущую частоту или применением линейного кодирования. Передача на поднесущей не является оптимальным выбором, поскольку ОСШ для приемного устройства с поднесущей в 4 раза меньше ОСШ, которое получается с помощью приемника прямого детектирования и прямой модуляции несущей. Наиболее приемлемым представляется использование линейного кодирования информационного сигнала перед оптической модуляцией, в частности применение линейных кодов вида 1B2B и mBnB [1]. Также может быть использован код RZ, который является де-факто стандартом линейного кодирования в магистральных волоконно-оптических системах передачи. Необходимо отметить, что при использовании кода RZ (как и 1B2B) ширина спектра передаваемого сигнала будет равна удвоенной битовой скорости передачи.

Имитационная модель волоконно-оптической системы передачи

Моделирование производилось с помощью программного пакета OptiSystem компании OptiWave Inc. OptiSystem – это пакет моделирования оптических систем телекоммуникаций, разработанный для проектирования, тестирования и оптимизации каналов практически любого типа на физическом уровне. Симулятор системного уровня, основанный на реалистичном моделировании волоконно-оптических систем телекоммуникаций, OptiSystem обладает мощной средой моделирования и иерархическим определением компонентов и систем. Его возможности можно легко расширить за счет добавления пользовательских компонентов. OptiSystem подходит для широкого спектра приложений, от проектирования сетей CATV/WDM и колец SONET/SDH до проектирования оптических передатчиков, каналов, усилителей и приемников. Он имеет большую базу активных и пассивных компонентов, от простых лазерных диодов и фотодиодов, волокон, разветвителей, аттенюаторов, модуляторов до комплексных передающих и приемных оптических модулей, волоконных усилителей, регенераторов и т.д. Также широко представлены измерительные устройства оптического и электрического диапазонов. Большой набор настраиваемых параметров позволяют пользователю мониторить и оптимизировать конкретные технические характеристики устройств для повышения производительности системы в целом [4].

Имитационная модель волоконно-оптической системы передачи с трансформацией спектра и применением усилителя Мальденштама-Бриллюэна приведена на рис. 1. Система состоит из трех частей: передатчика, приемника и среды передачи.

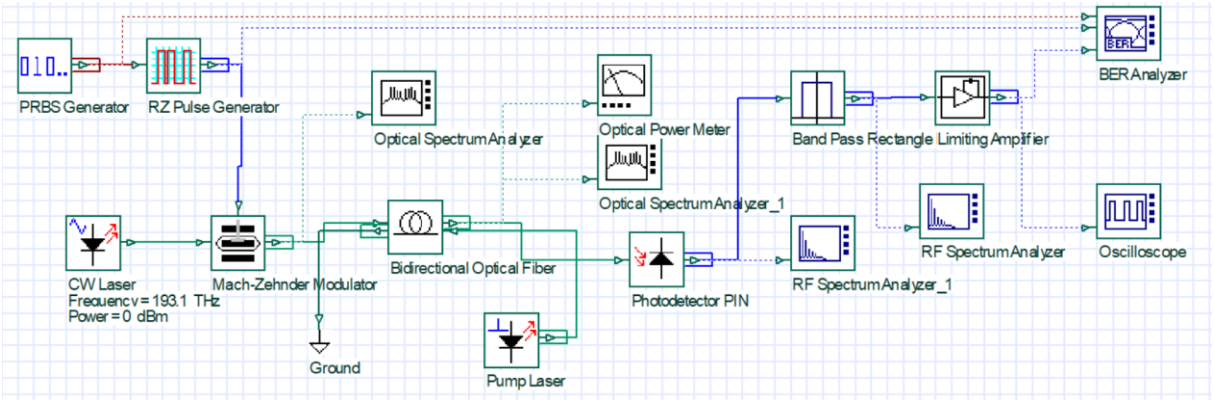


Рис. 1. Модель волоконно-оптической системы передачи

В качестве источника оптического излучения выступает полупроводниковый лазер, работающий в непрерывном режиме с частотой 193,1 ТГц (длина волны 1,55 мкм) и выходной мощностью 0 дБм. Передаваемые данные формируются генератором псевдослучайной последовательности, далее поступают на формирователь импульсов, где кодируются кодом RZ. Битовая скорость потока данных выбрана равной 2,5 Гбит/с. Это обусловлено тем, что частота лазера накачки распределенного усилителя Мандельштама-Бриллюэна должна отличаться от частоты несущего колебания на фиксированную величину, равную частоте смещения Бриллюэна, которая составляет примерно 11 ГГц для современных одномодовых волокон [5]. Если ширина спектра передаваемого сигнала превышает указанную величину, то излучение лазера накачки будет попадать в полосу сигнала, что приведет к существенным искажениям последнего. Несмотря на то, что излучение накачки и полезного сигнала распространяются в противоположных направлениях, влияние будет существенным за счет Рэлеевского обратного рассеяния и нелинейных эффектов в волокне. Для формирования линейного сигнала применен внешний модулятор Маха-Цандера, который производит модуляцию оптического излучения по интенсивности. Средой передачи является одномодовое оптическое волокно с затуханием 0,2 дБ/км. На приемной стороне установлен оптический усилитель Мандельштама-Бриллюэна, представляющий собой лазер накачки с частотой 193,111 ТГц, излучение которого вводится в оптическое волокно навстречу передаваемому сигналу. Усилительной средой является оптическая линия связи, поскольку эффект Бриллюэна возникает в стандартном волокне и не требует специальных волокон в отличие от усилителей EDFA. Оптическое изучение с усиленной

несущей детектируется p-i-n-фотодетектором. Электрический сигнал с выхода детектора проходит через полосовой фильтр. Применение именно полосового фильтра вместо ФНЧ обусловлено необходимостью подавления низкочастотных компонентов, которые подверглись существенному усилению вместе с несущей вследствие хоть и крайне узкой, однако ненулевой полосы усиления, и вызывают нелинейные искажения принятого сигнала. С целью более существенного подавления нелинейных искажений восстановленного цифрового сигнала применен усилитель-ограничитель. Для визуального контроля и измерения параметров передаваемого и применяемого сигнала использовались оптический и электрический анализаторы спектра, осциллограф, измеритель оптической мощности. Оценка эффективности системы в целом применялся анализатор битовых ошибок канала, позволяющий вычислять коэффициент битовых ошибок (BER), а также строить глаз-диаграмму принятого сигнала.

Результаты моделирования

Избирательное усиление оптической несущей можно наблюдать визуально в спектре оптического сигнала на входе фотодетектора, который получен при мощности лазера накачки 0,91 мВт и величине смещения Бриллюэна 11,0 ГГц (рис. 2).

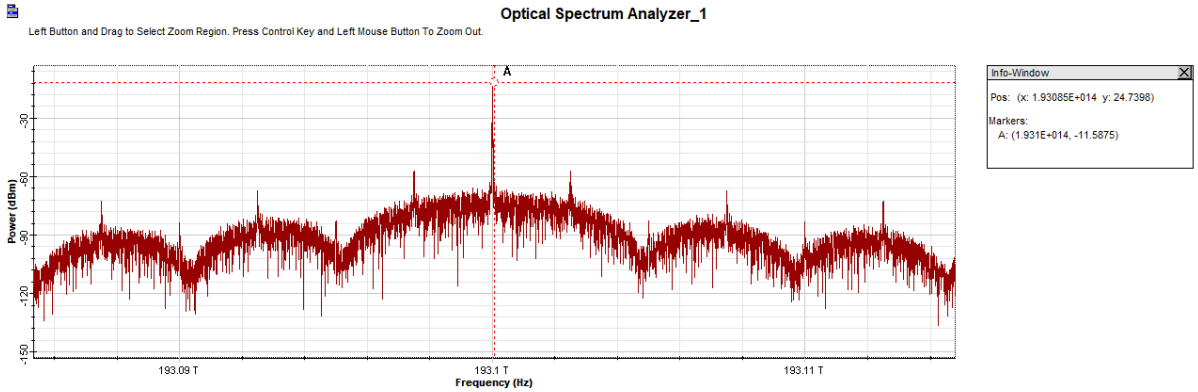


Рис. 2. Оптический спектр сигнала после избирательного усиления

Как можно заметить уровень несущей более чем на 45 дБ превосходит остальные компоненты спектра. Также усиление может быть оценено по измерению оптической мощности на выходе волокна. Зависимость оптической мощности на выходе волокна от мощности лазера накачки, полученная при величине смещения Бриллюэна 11 ГГц и мощности оптического сигнала без усиления –40,3 дБм, приведена на рис. 3.

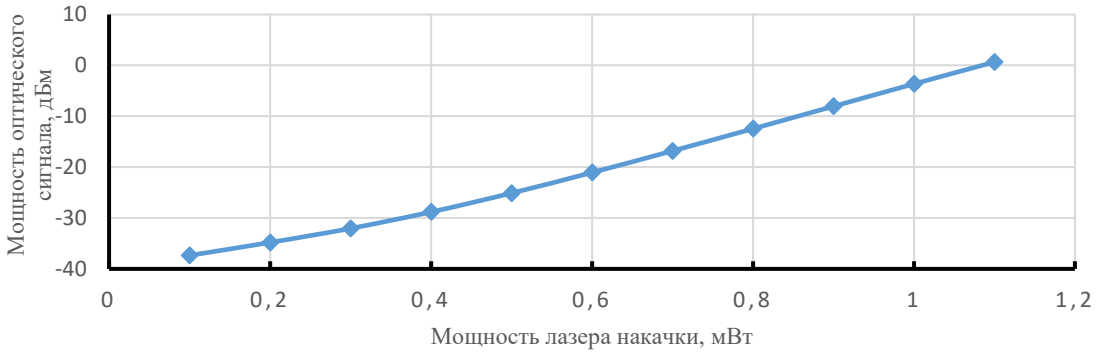


Рис. 3. Зависимость оптической мощности на выходе волокна от мощности лазера накачки

Как видно из графика, мощность сигнала линейно возрастает с ростом мощности лазера накачки. Однако, BER уменьшается с ростом усиления только до определенной мощности, а затем начинает увеличиваться, что может быть объяснено увеличением нелинейных искажений сигнала после детектирования. Зависимости BER и Q-фактора от мощности лазера накачки, полученные при тех же параметрах, приведены на рисунке 4. Можно заметить, что данные зависимости имеют выраженный экстремум, когда Q-фактор принимает свое максимальное

значение, а BER – минимальное. Таким образом можно произвести оптимальный выбор мощности лазера накачки, значение которой составит 0,91 мВт.

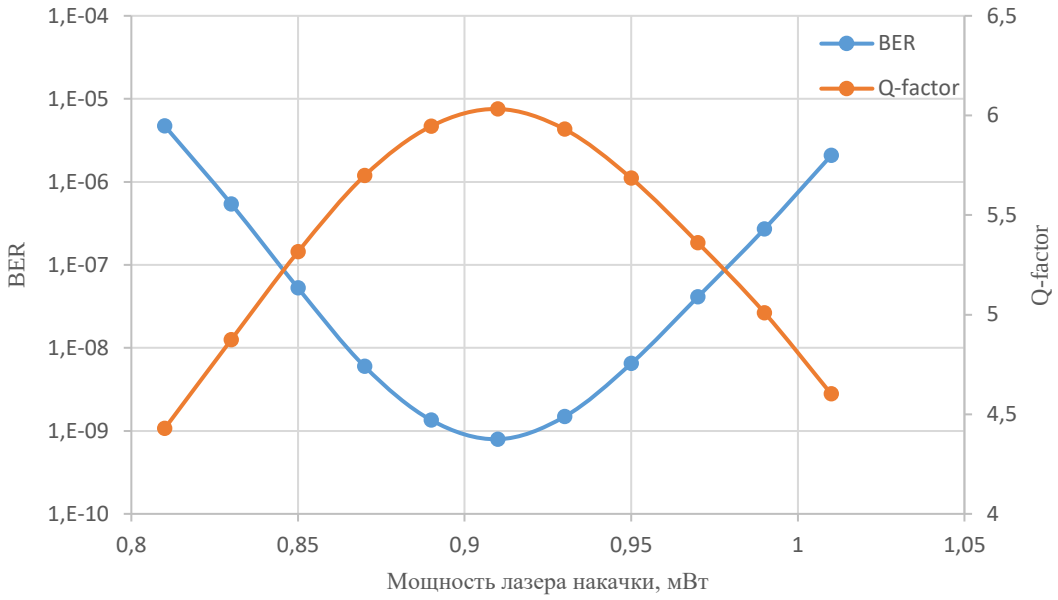


Рис. 4. Зависимость BER и Q-фактора от мощности лазера накачки

Длина оптического волокна и соответственно затухание в линии связи выбраны максимально возможными для обеспечения коэффициента битовых ошибок на более 10^{-9}. При такой длине волокна была измерена оптическая мощность на входе приемного модуля без усиления, которая составила –40,3 дБм. Эта величина является чувствительностью приемника, что показывает выигрыш по чувствительности исследуемого приемного модуля по сравнению с приемником без усилителя более чем на 10 дБм [6].

Также было произведено исследование зависимости коэффициента усиления усилителя от величины смещения Бриллюэна. График данной зависимости при мощности лазера накачки 0,91 мВт, длине оптического волокна 170 км и мощности оптического сигнала на выходе волокна –40,3 дБм приведена на рисунке 5. Наибольшее усиление было получено при величине смещения Бриллюэна 11 ГГц, что соответствует [5]. Максимальный коэффициент усиления при допустимом BER ($<10^{-9}$) составил 32,7 дБ.

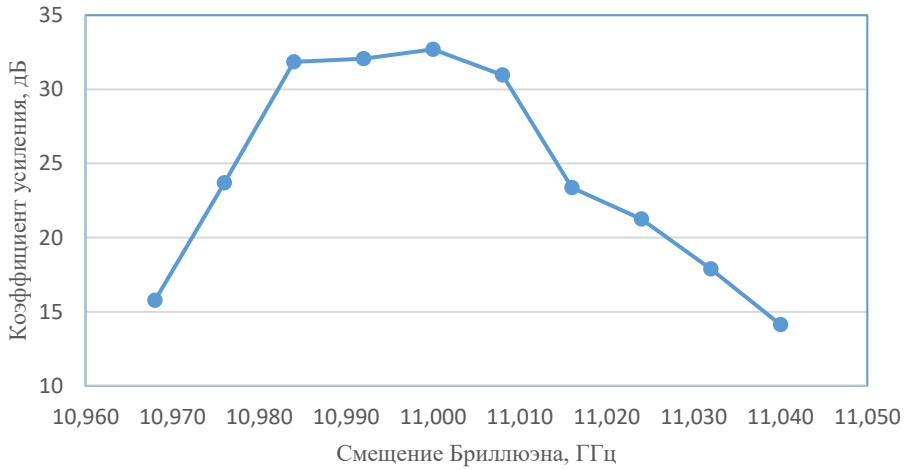


Рис. 5. Зависимость коэффициента усиления усилителя от величины смещения Бриллюэна

Существенную роль в восстановлении электрического сигнала и получении требуемого BER играет полосовая фильтрация. Были исследованы зависимости BER и Q-фактора

восстановленного сигнала от ширины полосы пропускания полосового фильтра, которая приведена на рис. 6. Видно, что максимальное значение Q-фактора и минимальное значение BER достигается при ширине полосы пропускания фильтра равной 2,4 ГГц.

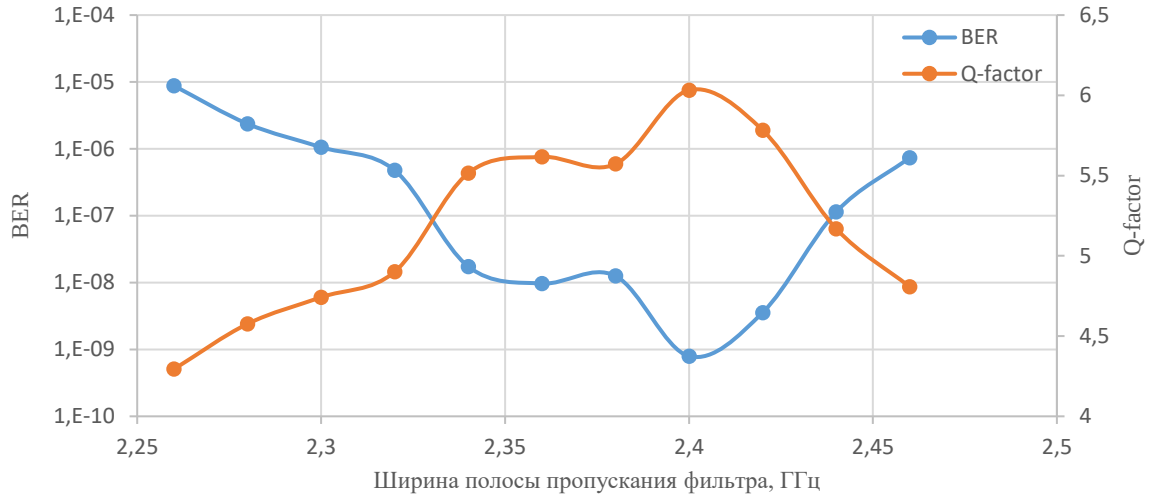


Рис. 6. Зависимость BER и Q-фактора от ширины полосы пропускания полосового фильтра

На рис. 7 приведена глаз-диаграмма электрического сигнала на выходе приемного оптического модуля и результат измерения BER при мощности лазера накачки 0,91 мВт, длине оптического волокна 170 км, мощности оптического сигнала на выходе волокна –40,3 дБм, величине смещения Бриллюэна 11 ГГц, коэффициенте усиления 32,7 дБ.

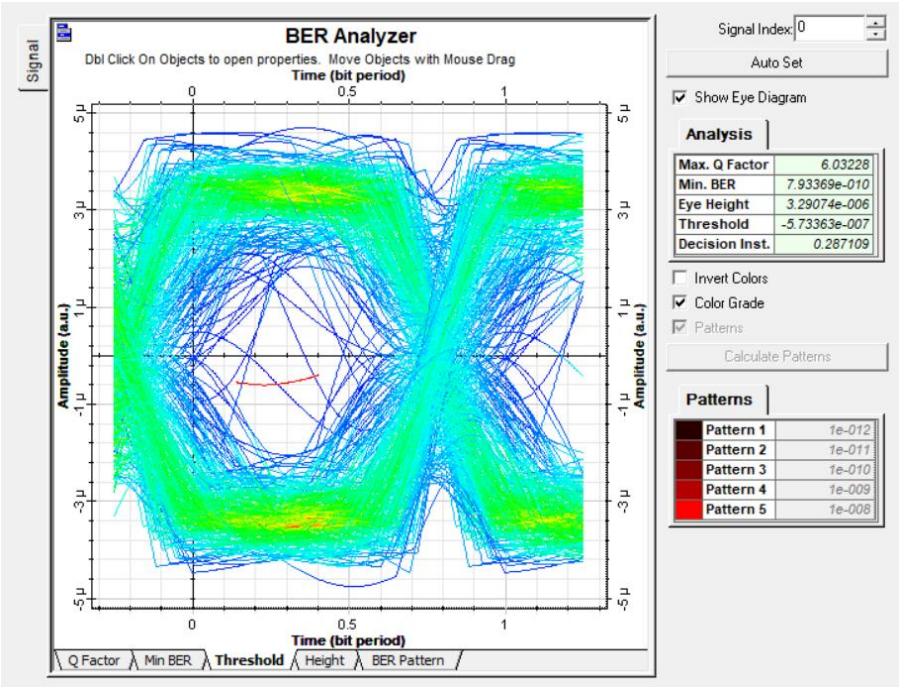


Рис. 7. Глаз-диаграмма электрического сигнала на выходе приемного оптического

Заключение

В результате моделирования волоконно-оптической системы передачи с избирательным усилением оптического несущего колебания и трансформацией спектра линейного сигнала получены следующие результаты. Определен оптимальный уровень мощности лазера накачки волоконно-оптического усилителя Мандельштама-Бриллюэна для обеспечения минимального BER, который составил 0,91 мВт. Показано, что коэффициент усиления при это составил 32,7 дБ,

что соответствует теории [2]. Получена чувствительность исследуемого приемного оптического модуля, равная –40,3 дБм, что соответствует теоретическим расчетам [1]. Разработана структура последетекторной части приемного модуля, которая отличается использованием полосового фильтра и усилителя-ограничителя.

OPTICAL CARRIER AMPLIFICATION TO IMPROVE THE RECEIVER EFFICIENCY IN FIBER-OPTIC COMMUNICATION SYSTEMS

Y.V. ROSHCHUPKIN

Abstract. A simulation of the fiber-optic communication system with increased efficiency of the optical receiver due to selective amplification of the optical carrier and the transmit signal spectrum transformation was carried out. The structure of the optical receiver module was substantiated and evaluated. The optimal parameters of the fiber-optic amplifier and post-detection processing circuits were determined.

Keywords: fiber-optic communication systems, optical receiver module, stimulated Brillouin scattering, optical amplifier.

Список литературы

1. Урядов В.Н., Рощупкин Я.В., Бунас В.Ю. и др. // Докл. БГУИР. 2015. № 8 (94). С. 11–16.
2. Yeniay A., Delavaux J.-M., Toulouse J. // J. Lightwave Techol. 2002. Vol. 20. № 8. P. 1425–1432.
3. Урядов В.Н., Рощупкин Я.В., Бунас В.Ю. и др. // Докл. БГУИР. 2016. № 4 (98). С. 10–14.
4. Khadir A. A., Dhahir B. F., Fu X. // International Journal of Computer Science and Mobile Computing. 2014. Vol. 3, Iss. 6, P. 42 – 53.
5. Фриман Р. Волоконно-оптический системы связи. Пер. с англ. под ред. Н. Н. Слепова. М.: Техносфера, 2003.
6. Урядов В.Н., Стункус Ю.Б. // Докл. БГУИР. 2006. № 3 (15). С. 48–53.

UDC 621.391.14

SYSTEM DESIGN OVERVIEW OF HAMMING PRODUCT CODES IMPLEMENTATION ON FPGA

Y.M. CHEN, X.H. REN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 06, 2023

Abstract. This paper describes the implementation of a Hamming product codes encoder and decoder on a FPGA. The encoder and decoder design, including hardware structure and logic design.

Keywords: hardware design, Hamming product codes, FPGA, logic design.

Introduction

Based on Hamming product codes theory and using FPGA as the hardware platform, this project aims to provide a reliable and efficient data transmission solution. The main application area is on-chip interconnection or on-board interconnection, which can realize fast and reliable data transmission, thus improving the efficiency and reliability of data transmission. The main functions of the project include coding, error detection, error correction and hybrid re-request, which can effectively reduce the risk of transmission errors and lost data and ensure data integrity and reliability.

In this project, encoding, error detection and error correction are the key basic functions. Encoding converts the original data into Hamming product codes to improve the reliability of transmitted data, while error checking and error correction are performed at the receiving end to ensure the integrity of the data. This project also has a hybrid re-request function, which automatically requests data retransmission when an error occurs in data transmission, thus reducing transmission loss and improving data transmission reliability.

This project also implements the following functional features: low latency transmission through segmented transmission, which divides data into smaller parts, thereby increasing transmission speed and reducing transmission latency. The efficiency of data transmission can be further improved by increasing the clock frequency of the transmission module through clock partitioning. By utilizing the parallel computing capability of FPGAs, lower coding latency and better error detection and correction rates are achieved. These features allow the project to adapt to different application scenarios and provide an efficient and reliable solution for various data transmission needs.

In the fields of on-chip interconnect and on-board interconnect; the efficiency and reliability of data transmission are critical. The implementation of this project can provide reliable and efficient data transmission guarantee, thus contributing to the development and progress of these fields. In addition, the project can be widely used in communication, data storage, intelligent manufacturing, etc., providing efficient and reliable data transmission solutions for these fields and promoting the development and innovation in these fields.

Product codes

Product codes are serially concatenated codes, which were presented by Elias in 1954 [1]. The construction method of product codes allows us to construct long, powerful codes from short assembly codes. As shown in Figure 4, it is a schematic diagram of the code word structure of a two-dimensional Hamming product codes. The coding method uses a binary linear code $C(n,k)$, where the code word length is *n*, the message length is *k*, and the code distance is *d*. The product codes C_p of this coding

method can be expressed as $C(n,k)\times C(n,k)=C_p(n^2,k^2)$, where the code word length of the product codes is n^2, the message length is k^2, and the code distance is d^2 [2].

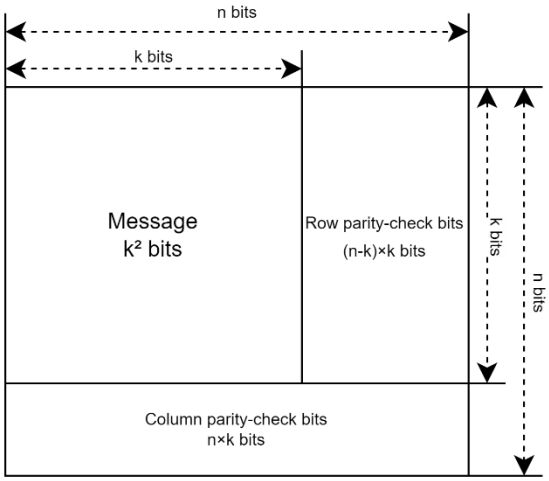


Figure 4. Codeword structure diagram

Specifically, the *k* messages of length *k* are first arranged into a matrix of *k* rows and *k* columns in a row-first order. Then each row in the matrix is encoded using binary linear coding *C* to generate *k* row parity bits of length $n-k$, and these parity bits are appended to the end of the message matrix to obtain a matrix of *k* rows and *n* columns. Next, consider this matrix as a new message matrix and encode each column in it, again using *C* for encoding, to generate n columns of column parity bits of length $n-k$. Finally, these column parity bits are appended to the end of the matrix to obtain a code word of length n^2.

System design overview

As shown in Figure 5 below, the design is mainly based on the message transmission direction and is divided into two main functional parts: transmitting and receiving. Implementing the code for the part of the virtual line is the main goal of this project [3,4].

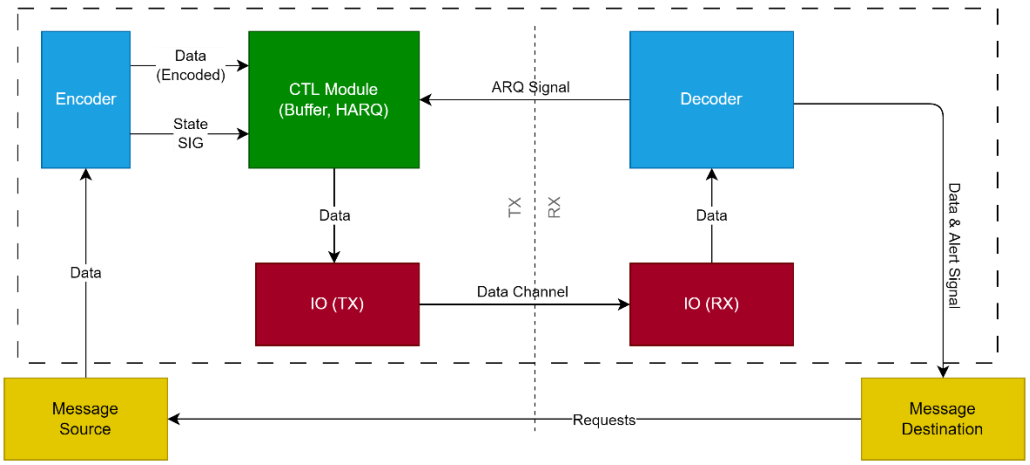


Figure 5. System overall structure diagram

Among them, the high-speed IO port is a general-purpose component for sending and receiving messages. The components unique to the sender include the message source, the encoder, and the control module. The source is responsible for generating the message to be sent, the encoder will encode the message for errors caused by the transmission, and the control module is responsible for managing the process of sending the message. In contrast, the only components that are unique to the receiver are the

message destination and the decoder. The message destination is the final point of reception of the message, and the decoder will verify the received message and correct errors if feasible. With this design, the sender can send the message to the receiver in a reliable and efficient manner. In addition, since the high-speed IO port is a very flexible and versatile component, it can be easily adapted to different clock frequencies and bit widths, increasing the scalability and flexibility of the system.

The control module of the sender is an important part of the whole system, which is not only responsible for managing the message sending process, but also enables the function of hybrid re-request. Specifically, it can cache the encoded messages and wait for the acknowledgement signal from the receiver. If the receiver fails to acknowledge in time or returns an error code, the control module sends supplementary bits or retransmits the message according to the set logic until the receiver acknowledges. This hybrid re-request mechanism can ensure transmission reliability while minimizing the number of retransmissions and improving the efficiency of the system.

The high-speed IO port serves as a clock demarcation in the system, and it can adapt higher frequency clock sources to achieve high-speed transmission. When the high-speed IO port receives a message from the sender, it transmits the message to the receiver at a high bit rate. To ensure the accuracy and reliability of the transmission, the high-speed IO port must be synchronized. By using this IO port, the whole system can achieve high-speed transmission to meet the high-bandwidth and low-latency transmission requirements.

Encoder design and implementation

The structure diagram of the encoder is shown in Figure 6, which includes four main parts: IO, control, storage (independent input-output buffer), and calculation units. The main logic control of the encoder is implemented by FSM. The calculation unit is composed of multiple parallel calculation units, whose working state is controlled by FSM state and feedback to the control unit.

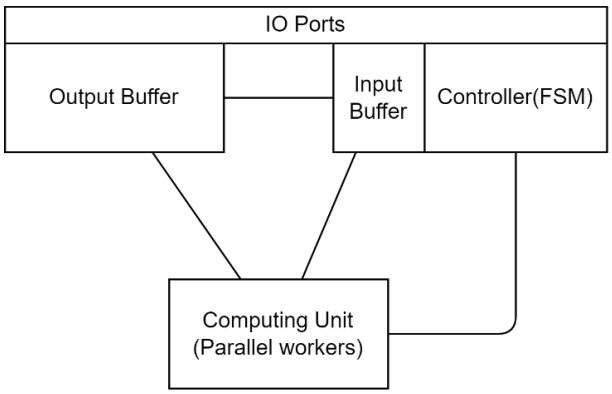


Figure 6. Encoder structure block diagram

The output of the encoder is directly transmitted to the output buffer. The control unit in this design also includes a module for implementing array transpose, which is used to transpose the data in the input buffer. By doing so, the encoder can achieve two-step encoding, thereby improving encoding efficiency.

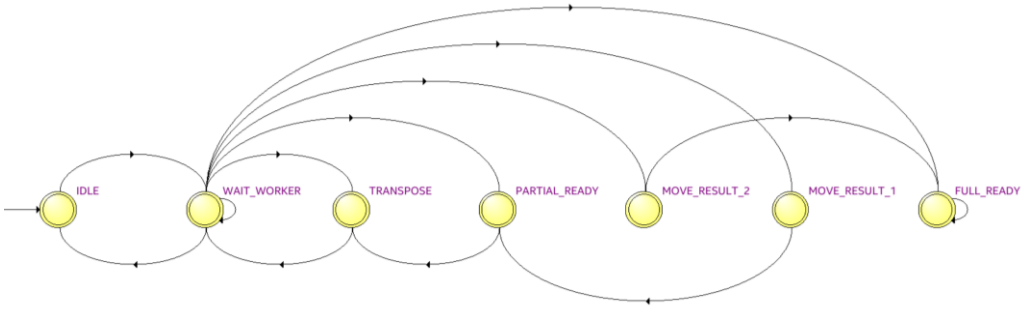


Figure 7. State transfer diagram of encoder FSM

The core of the encoder is its control unit, whose main functional logic is controlled by its internal finite state machine, as shown in Figure 7, which has the following seven states: IDLE, READY, PARTIAL_READY, WAITING_WORKER, transpose for second encoding stage (TRANSPOSE), and two states for internal data transfer (MOVE_RESULT_1 and MOVE_RESULT_2).

The operating states of the parallel computing unit are managed by its internal controller, which communicates with the encoder control unit in the form of a signal. In order to reduce the delay caused by the two-step encoding, the encoder enters the "partially completed" state after completing the row parity bit operation and can provide a code word without column parity bits for transmission.

Decoder design

The block diagram of the decoder is shown in Figure 8 below. The decoder of this project contains HDL implementations of several algorithms, so it also contains an additional logic module or set of logic modules that are responsible for adapting different algorithmic logic, depending on the construction parameters, compared to the encoder.

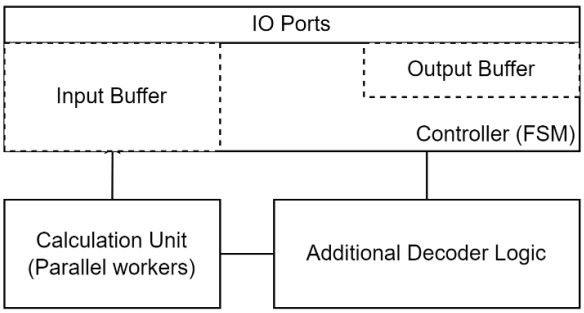


Figure 8. Decoder core structure block diagram

Similar as the encoder, the control of the decoder is controlled by a finite state machine, but its logic is more complex due to the addition of the auto-request logic. As shown in the Figure 9 below, this finite state machine includes: standby (IDLE), waiting for complete codeword (WAIT_FULL_CODE), two ongoing checks (CHK_PARTIAL, CHK_FULL) and two ready states (RDY_OK, RDY_FAIL). The first state of this finite state machine is the standby state (IDLE), where the decoder is idle and waiting for input. If all the row parity checks bits available, the decoder transitions to the check for partial codewords state (CHK_PARTIAL). If the check for partial codewords fails, the decoder transitions to a waiting state (WAIT_FULL_CODE) and waits for the missing data (column parity check bits) to arrive. Once a whole codeword available, the decoder transitions to the check for complete codewords state (CHK_FULL), the check is running in a sepreate process, which is achived by an entity installation controlled by generic mapping and generate. When the any of the tow check is successful, the decoder enters the ready state (RDY_OK), indicating that the data is not corrupted and can output it. If the checks fail, the decoder enters the ready failure state (RDY_FAIL), indicating that the data is corrupted, and the decoder cannot correct it and must wait for input again.

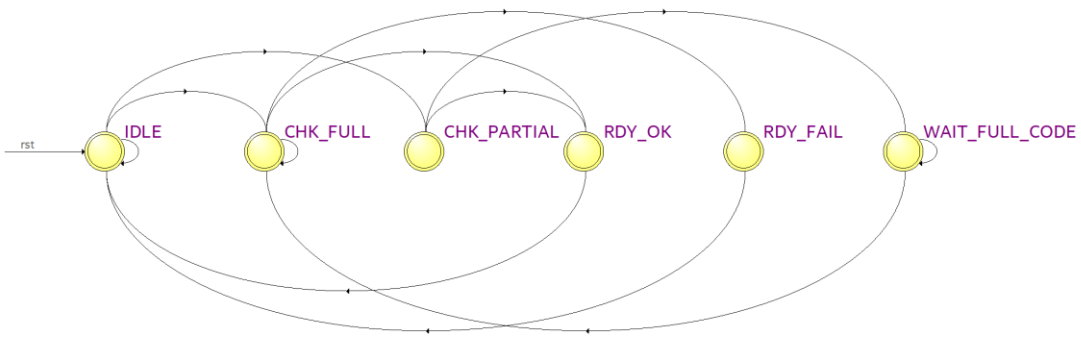


Figure 9. State transition diagram of the decoder FSM

Build and test system

The goal of this project is to implement a small communication system on FPGA in HDL that includes a transmitter, receiver, message source, message destination, and transmission channel. In order to achieve the above functionality, the project required a complete and easy-to-use tool chain for building and testing.

In the spirit of open source first, VSCode is used as the code editor, Gradle as the build environment, GHDL as the syntax checking, elaborate and emulate tool, and TypeScript, Python and GNU Octave as the auxiliary code generation tools. The project also provides a pre-built development environment in a Docker image, and a "devcontainer" configuration file to optimize development efficiency. To simplify the workflow of writing, compiling, and testing, a dedicated build script and a Gradle command wrapper were written.

Conclusion

In conclusion, this paper provides a brief description of the implementation of a Hamming product code encoder and decoder based on FPGA. Firstly, the basic principles of the algorithm are introduced, followed by a detailed discussion on the design of the encoder and decoder, including hardware structure and logic design. Then, the paper describes the building and testing system framework of the project. Through testing, the implemented encoder and decoder on FPGA demonstrate excellent performance, meeting the requirements of high-speed data transmission and low power consumption. This implementation provides a feasible solution for the design of FPGA-based encoder and decoder, with high practicality and reliability.

References

1. Elias P. // In Transactions of the IRE Professional Group on Information Theory. 1954. Vol. 4. P. 29–37.
2. Lin. S., Costello D.J. Error Control Coding, 2nd edition. USA, 2004.
3. Bo F., Ampadu P. // 2008 IEEE International SOC Conference. 2008. P. 59–62.
4. 毕涛., 刘迪., 张大为., 葛宝川. // 现代信息科技. 2023. P.58–63.

УДК 004.93'1

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ БИБЛИОТЕК РАСПОЗНАВАНИЯ ЛИЦ

С.Н. ПЕТРОВ[1], А.Д. МАТЮШЕНКО[2], Д.А. РУДЕНЯ[2]

1 – Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь
2 – Национальный детский технопарк, Республика Беларусь

Поступила в редакцию 20 марта февраля 2023

Аннотация. Приведены результаты сравнения эффективности библиотек распознавания лиц Face-Recognition, Dlib, DeepFace на изображениях базы Labeled Faces in the Wild (LFW). Показано, что библиотека Face-Recognition является лучшим выбором в сравнении с Dlib и DeepFace.

Ключевые слова: распознавание лиц, сверточные нейронные сети, биометрия, Face-Recognition, Dlib, DeepFace.

Введение

Распознавание лиц, это технология, которая позволяет автоматически идентифицировать или верифицировать личность человека по его изображению. Распознавание человека по изображению лица имеет ряд преимуществ по сравнению с другими методами, а именно [1], не требует специального дорогостоящего оборудования для реализации и является бесконтактным. Для большинства приложений достаточно видеокамеры для получения изображения и компьютера для его обработки. Результаты многих исследований показали [2, 3], что сверточные нейронные сети являются эффективным средством для распознавания лиц. Соответственно, появилось множество фреймворков и библиотек для создания и тренировки нейронных сетей под различные задачи. В работе проводится сравнительный анализ эффективности популярных библиотек распознавания лиц Face-Recognition, Dlib, DeepFace.

Библиотеки распознавания лиц

Библиотека Face-Recognition, основанная на библиотеке Dlib, позволяет распознавать лица на изображениях и в потоке видео. Основная особенность этой библиотеки – это ее скорость и точность. Face-Recognition использует множество методов для обработки изображений и выделения лиц, включая поиск лиц по признакам, детектирование ключевых точек на лицах, а также использование глубоких нейронных сетей для классификации лиц. Благодаря этим методам, Face-Recognition обеспечивает высокую точность распознавания, а также высокую производительность.

Библиотека Dlib также широко используется для распознавания лиц и имеет ряд преимуществ. Она использует современные методы машинного обучения, включая SVM-классификаторы и глубокие нейронные сети. Кроме того, она имеет свой собственный формат файла для хранения обученных моделей, что делает ее более удобной для использования.

Библиотека DeepFace является еще более новой и использует глубокие нейронные сети для распознавания лиц. Она предоставляет набор предобученных моделей для классификации лиц, которые можно использовать для быстрого и точного распознавания. Однако, из-за сложности и объемности глубоких нейронных сетей, производительность DeepFace может быть значительно ниже, чем у более легковесных библиотек.

Рассмотрим подробнее методы, используемые в библиотеках для распознавания лиц.

В библиотеке Dlib для распознавания лиц используется модель Dlib_face_recognition_resnet_model_v1. Эта модель использует глубокие нейронные сети для извлечения высокоуровневых признаков лиц, таких как расположение глаз, носа и рта. Затем, на основе этих признаков, модель генерирует уникальный вектор, называемый «вектором признаков», который представляет собой компактное числовое представление лица. Для сравнения лиц в Dlib используется евклидово расстояние между векторами признаков. Чем меньше расстояние между двумя векторами, тем больше вероятность, что они соответствуют одному и тому же лицу.

В библиотеке Face-Recognition для сравнения лиц используется метод compare_faces. Этот метод также использует вектора признаков, сгенерированные с помощью глубоких нейронных сетей. Он принимает два вектора в качестве входных данных и возвращает значение True или False, в зависимости от того, соответствуют ли они одному и тому же лицу. Для сравнения лиц в методе compare_faces также используется евклидово расстояние между векторами признаков.

В библиотеке DeepFace для верификации лиц используется метод verify. Этот метод принимает два изображения и определяет, соответствуют ли они одному и тому же лицу. Для этого он использует сверточные нейронные сети, предобученные на большом наборе данных лиц. Он генерирует векторы признаков для каждого изображения и затем вычисляет расстояние между ними. Если расстояние меньше заданного порога, метод verify вернет значение True, что означает, что изображения соответствуют одному и тому же лицу.

Таким образом, евклидово расстояние между векторами признаков является ключевым элементом для сравнения и верификации лиц в библиотеках для распознавания лиц. Оно позволяет быстро и точно определять, соответствуют ли два изображения одному и тому же лицу, основываясь на вычислении расстояния между векторами. Благодаря этому подходу, эти библиотеки показывают высокую точность распознавания лиц.

Для исследования эффективности указанных библиотек была использована база данных фотографий лиц Labeled Faces in the Wild (LFW) [4]. Эта база содержит 13233 изображения 5749 человек, которые были собраны в Интернете. Каждое изображение имеет размер 250×250 пикселей в формате jpg.

Результаты исследования эффективности библиотек

Код, написанный на языке Python для реализации каждого из методов, продемонстрирован на рис. 1–3.

```
import os
import shutil
import face_recognition

dir_list = os.listdir("C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled")
T = 0
N = 0
R = 0

for folder_name in dir_list:
    file_names = os.listdir(
        f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled/{folder_name}")
    R += 1
    if R % 100 == 0:
        print(T / N)
        print(str(R) + "_____________________________")
    if len(file_names) >= 2:
        N += 1
        pic1 = face_recognition.load_image_file(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[0]}")
        try:
            pic1_encode = face_recognition.face_encodings(pic1)[0]
        except IndexError:
            continue
        pic2 = face_recognition.load_image_file(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[1]}")
        try:
            pic2_encode = face_recognition.face_encodings(pic2)[0]
        except IndexError:
            continue
        results = face_recognition.compare_faces([pic1_encode], pic2_encode)
        print(N)
        if results[0] == True:
            T += 1
print(T / N)
```

Рис. 1. Код для тестирования метода распознавания лиц на основе библиотеки Face-Recognition

```python
from deepface import DeepFace
import cv2
import os
import shutil

dir_list = os.listdir("C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled")
T = 0
N = 0
R = 0

for folder_name in dir_list:
    file_names = os.listdir(
        f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled/{folder_name}")
    R += 1
    if R % 100 == 0:
        print(T / N)
        print(str(R) + "_____________________________")
    if len(file_names) >= 2:
        N += 1
        img1 = cv2.imread(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[0]}")
        img2 = cv2.imread(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[1]}")
        try:
            results = DeepFace.verify(img1, img2)
        except ValueError:
            continue
        print(N)
        if results['distance'] < 0.6:
            T += 1
print(T / N)
```

Рис. 2. Код для тестирования метода распознавания лиц на основе библиотеки DeepFace

```python
import dlib
import numpy as np
import os
import shutil
import face_recognition

dir_list = os.listdir("C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled")
T = 0
N = 0
R = 0
detector = dlib.get_frontal_face_detector()
sp = dlib.shape_predictor("shape_predictor_5_face_landmarks.dat")
facerec = dlib.face_recognition_model_v1("dlib_face_recognition_resnet_model_v1.dat")
for folder_name in dir_list:
    file_names = os.listdir(
        f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-deepfunneled/{folder_name}")
    R += 1
    if R % 100 == 0:
        print(T / N)
        print(str(R) + "_____________________________")
    if len(file_names) >= 2:
        N += 1
        img1 = dlib.load_rgb_image(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[0]}")
        img2 = dlib.load_rgb_image(
            f"C:/Users/fluffy11/PycharmProjects/pythonProject3/reco_test/lfw-
deepfunneled/{folder_name}/{file_names[1]}")
        img1_detection = detector(img1, 1)
        img2_detection = detector(img2, 1)
        try:
            img1_shape = sp(img1, img1_detection[0])
            img2_shape = sp(img2, img2_detection[0])
        except IndexError:
            continue

        img1_aligned = dlib.get_face_chip(img1, img1_shape)
        img2_aligned = dlib.get_face_chip(img2, img2_shape)
        img1_representation = facerec.compute_face_descriptor(img1_aligned)
        img2_representation = facerec.compute_face_descriptor(img2_aligned)
        img1_representation = np.array(img1_representation)
        img2_representation = np.array(img2_representation)
        euclidean_distance = img1_representation - img2_representation
        euclidean_distance = np.sum(np.multiply(euclidean_distance, euclidean_distance))
        euclidean_distance = np.sqrt(euclidean_distance)

        print(N)
        if euclidean_distance < 0.6:
            T += 1
print(T / N)
```

Рис. 3. Код для тестирования метода распознавания лиц на основе библиотеки Dlib

Результат тестирования библиотек распознавания лиц представлен в табл. 1., рассчитана метрика «точность» (Accuracy).

Табл. 1. **Результаты сравнения эффективности библиотек распознавания лиц**

Название библиотеки	Accuracy
Face-Recognition	96,13%
DeepFace	93,63%
Dlib	96,07%

Как видно из таблицы, библиотека Face-Recognition показывает большую точность среди рассмотренных библиотек. Реализация на Dlib также показала высокую точность, но при этом программный код для распознавания оказался более массивным.

Заключение

Согласно результатам исследования методов распознавания лиц, отметим, что Face-Recognition является лучшим выбором в сравнении с Dlib и DeepFace. Причина этому заключается в том, что Face-Recognition использует оптимизированные алгоритмы, которые позволяют ей работать быстрее и эффективнее, чем другие библиотеки. Кроме того, Face-Recognition поддерживает не только распознавание лиц, но и обнаружение, а также выравнивание лиц, что делает ее полноценным инструментом для работы с изображениями лиц.

В целом, использование библиотек для распознавания лиц является важным инструментом для автоматической идентификации людей на изображениях и в видеопотоках. Благодаря высокой точности и быстродействию, эти библиотеки могут быть использованы в различных областях, таких как безопасность, медицина, социальные сети и т.д.

STUDY OF THE EFFECTIVENESS OF FACE RECOGNITION LIBRARIES

S.N. PETROV, A.D. MATSIUSHENKA, D.A. RUDENYA

Abstract. The results of comparing the effectiveness of face recognition libraries Face-Recognition, Dlib, Deep Face on the images of the Labelled Faces in the Wild (LFW) database are presented. It is shown that the Face-Recognition library is the best choice in comparison with Dlib and DeepFace.

Keywords: face recognition, convolutional neural networks, biometrics, Face-Recognition, Dlib, DeepFace.

Список литературы

1. Антончик, А.В. // Докл. БГУИР. 2009. № 2 (40). С. 67–72.
2. Мищенкова Е.С. // Вестник Волгоградского государственного университета. Серия 9. Исследования молодых ученых. 2016. № 11. С. 74–76.
3. Паршин С.Е. // Сборник научных трудов НГТУ. 2019. № 1 (94). С. 55–70.
4. Labelled Faces in the Wild (LFW) Dataset [Электронный ресурс]. URL: https://www.kaggle.com/datasets/jessicali9530/lfw-dataset.

UDC [004.732+004.432+004.738.5]-027.31

LAN INSTANT MESSAGE COMMUNICATION APPLICATION DESIGN BASED ON TCP

ZHANG BOWEN, ZHANG RONGLIANG, N.V. KHAJYNAVA

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 19, 2023

Abstract. The use of LAN Message Communication through TCP/IP provides a reliable, secure, and cost-effective means of communication among company or organizational staff members. This study includes the added functionality of file and picture transfers, which addresses issues related to time, cost, and information safety. The protocol offers a range of communication options, facilitating information exchange among individuals. Developed as a standalone application using C++ programming language and QT platform, the proposed protocol has been tested in the LANs of our institute.

Keywords: LAN, communication, TCP, UDP, IP.

Introduction

The evolution of communication technology has revolutionized the way we interact with each other. From the traditional modes of communication such as mail, telephone, and telegraph to the internet and digital convergence, communication has come a long way. The introduction of the internet has given rise to various means of communication, including instant messaging. In this context, the focus of this essay is on the necessity and design of a LAN instant message application based on TCP and C++. The need for such an application arises from the growing concern of data and message safety during transmission.

While conventional means of communication are still prevalent in many organizations, the advances in electronic messaging services and their synchronous nature can bring about different advantages, including improving organization activities, saving costs, energy, and safety of information. In this paper, we will explore the development of a LAN instant message application based on TCP and C++, discussing its features, advantages, and implementation. We will also examine the risks of information leakage and the necessity of such an application in today's digital age. The paper will be structured as follows: Section 1 will provide an introduction of the topic, discussing the need for secure communication tools. Section 2 of this research about the structure design of the client and server. Section 3 describes how the client connected to server. Section 4 is the project's protocol. Section 5 gives the output of project's implantation. Finally, it is the conclusion of the research or the project [1].

The reason – In today's interconnected world, communication is essential for the smooth functioning of any organization. However, with the increasing amount of data and messages being transmitted, the risk of information leakage has also grown exponentially. In recent times, we have witnessed several high-profile cases of data breaches, where sensitive information has been compromised, resulting in significant losses for companies and individuals.

For instance, in 2017, Equifax, one of the largest credit reporting agencies in the US, suffered a data breach that exposed the personal information of 147 million people. This breach was caused by a vulnerability in the company's web application, which hackers exploited to gain unauthorized access to sensitive data. The breach resulted in the loss of sensitive personal and financial information, including names, social security numbers, and credit card numbers. Similarly, in 2020, the video conferencing app, Zoom, faced severe backlash over privacy and security concerns. The app's popularity skyrocketed due to the COVID-19 pandemic, but it soon became evident that the app was not adequately secured,

and data breaches were rampant. Zoom's user data was sold on the dark web, and users' personal information was leaked, leading to severe consequences for individuals and companies alike. In such a scenario, the need for secure communication tools cannot be overemphasized. This is where a LAN instant message application based on TCP using C++ comes into play. Such an application can help organizations communicate securely within their internal network, minimizing the risk of data breaches and information leakage.

In the wake of this, it is an urge to use a safer communication way to transmit information. That is the reason why we design this LAN instant message communication application.

Server-Client Structure Design

To realize the program, Client/Server model is adapted. In the client/server programming model, a server program awaits and fulfills requests from client programs, which might be running in the same or other computers.

In this structure (Figure 1), server must be running first and waiting (listening) for a connection, server listening at specific port number and create thread for each client connection. Each client has the same port number that the server listening at it and hostname of server (or IP).

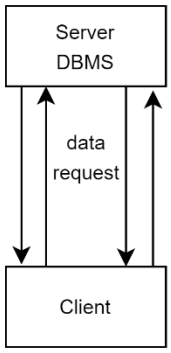


Figure 1. Structure of C/S

However, technically in the Client/Server model, a client cannot reach other clients directly. In other words, a client cannot send data to other clients directly. As Figure1 shown, a client can send data to another client or clients by send the request to server and then server send the data to the destination or target clients. How this information transmission works between server and client is based on TCP socket. Once the connection is established, data can be transmitted.

Another question arise – how do server and clients know what kind of information or action should server to respond to clients. That is the key point of this structure. Every Clients connected to the server have a matched socket to server and every operation made by clients will send an encode to server. Then server will get the request and give a corresponding response.

For instance, a client sends a message to server. The client sends an encode pattern like "@sendPersonalMesssage#ToFriendOne#content to server". Then the server will according the prefix of the encode "@sendPersonalMessage" to store the message in the database and simultaneously send an encode pattern like "@sendPersonalMessage#From#FriendTwo#content" to the target client. With the help of this encode pattern, TCP and Signal-emit from C++, communication is accomplished perfectly.

Database Design

As shown at the Figure 2, "many to many" relation is between entity user and group, which means a user can in many groups while a group can have many users.

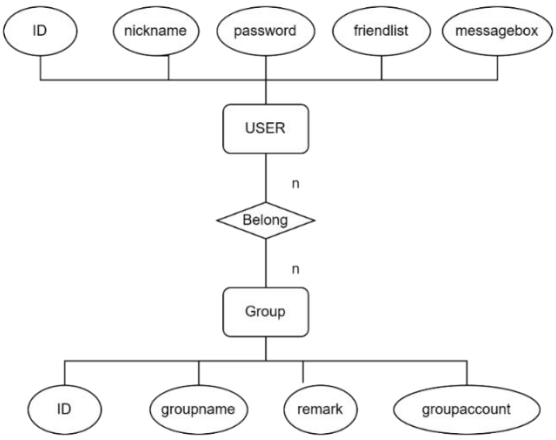


Figure 2. User and group

As shown at the Figure 3, “one to many” relation is between entity user and friend, which means a user, can have many friends. Each user has a friend table for its own and it has all friends contact.

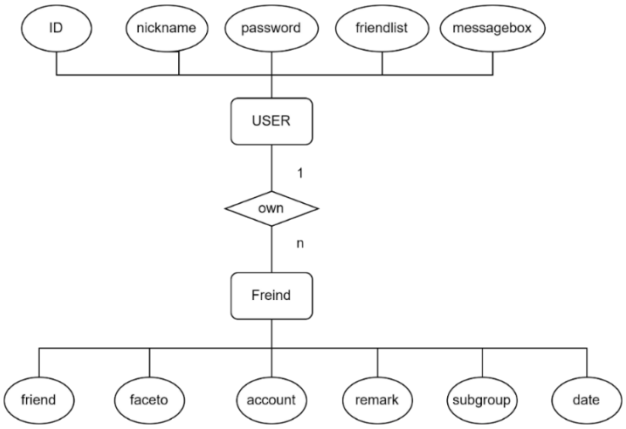


Figure 3. User and friend

As shown at the Figure 4, “one to many” relation is between entity user and message, which means a user, can have many messages from friends or groups. Each user has a messagebox table for its own, it has all messages from friends, and groups send to the user.

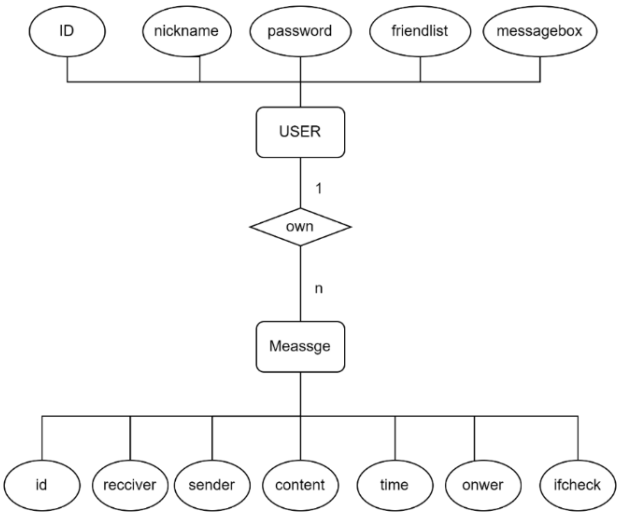


Figure 4. User and message

Comparative analysis of TCP and UDP

Two Internet Protocols can be chosen: TCP and UDP. After the comparative analysis, the TCP is used to implement data transmission.

TCP is connection-oriented while UDP is connectionless. For TCP, the client sends a synchronization request, the server sends back an acknowledgment, and the client returns a synchronization acknowledgment in response. Comparatively, for UDP, the message is sent out, without as much regard for the recipient, without considering the destination. Connectionless transport protocols can lose a minimal number of packets.

TCP sends data in a particular sequence, whereas there is no fixed order for UDP protocol. TCP send and receive data with sequence numbers. This allows the system to track the specific order in which data is transmitted, maintaining the desired sequence. On the contrary, UDP does not follow a sequencing mechanism. The protocol has no way of telling which data packets should come first, and if they are received in the wrong order. UDP also drops any data packet that it is unable to process. However, UDP is faster and more efficient than TCP User datagram protocol does not need an established connection to start sending packets. Therefore, it saves the time typically required to turn on the server and place it in a "passive open," listening state. It allows data transmission to begin faster immediately or extended latency time. There is also no need to put the packets in sequence or send and receive acknowledgments, saving time. UDP is also more efficient in terms of bandwidth. Once the data is in motion from the server to the client, TCP engages in many error check mechanisms, acknowledgment processes, and sequencing measures, which occupy a lot of bandwidth. In contrast, UDP quickly gets the data stream from one computing location to another without a lot of checks and balances. UDP is suitable for live and real-time data transmission, which TCP cannot support.

Despite its inherently unreliable nature, UDP continues to be a staple for online operations. This is because it is ideal for real-time data transmissions, where the loss of a few packets does not matter.

TCP is best for use cases where data integrity matters more than transmission speed. It will ensure that files and web pages arrive intact and can even be helpful for real-time analytics and content delivery networks, where dropped packets would fudge the outcomes. In comparison, UDP is suitable for media transmissions, such as video calling and online gaming [2].

Through what has been mentioned above, TCP is more reliable but less efficient and slower in transmitting. For Lan Instant Message Communication application, files and chat messages are the content that being transmitted. The transmission is not strictly real-time. Data integrity should come first. Therefore, it considers more reliability than efficiency. In this paper, TCP is used for the Internet Protocol of the application.

Conclusion

The proposed TCP protocol solved the instant communication problems among staff members and reduces time of communication at low cost. Simultaneously the LAN instant message communication application can guarantee the safety of the data transmission since it is not connected to the network. However, data leakage is still at risk, since Information leakage channels exist in any information space. A leakage channel in the most general sense is understood as an uncontrolled way of transmitting information. As a result, an attacker can gain unauthorized access to the confidential company data he needs. However, physical access by plugging unknown external devices still may cause a leakage. Therefore, organization application regulations should be combined with the application to reduce the risk of leakage.

References

1. Design and implement chat program using TCP/IP Mohammed A. Ahmed, Sara Ammar Rafea, Lara Moufaq Falah, Liqaa Samir Abd Ullah

2. TCP and UDP: Understanding 10 Key Differences Chiradeep BasuMallick [Electronic resource]. URL: https://www.spiceworks.com/tech/networking/articles/tcp-vs-udp/#_003

УДК 061.68

ПОТОКОВАЯ МОДЕЛЬ VPN С УЧЕТОМ ЗАДЕРЖКИ ПЕРЕДАЧИ ПАКЕТА В СЕТИ ЭЛЕКТРОСВЯЗИ СПЕЦИАЛЬНОГО НАЗНАЧЕНИЯ

С.С. ВРУБЛЕВСКИЙ, А.А. БЫСОВ

Военная академия Республики Беларусь, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. В статье представлена потоковая модель VPN, учитывающая задержку передачи пакета, при планировании VPN-туннелей в сети электросвязи специального назначения.

Ключевые слова: сеть электросвязи специального назначения, потоковая модель VPN, задержка передачи пакета, качество обслуживания пользователей.

Введение

Современная технологическая революция в области инфокоммуникаций, связанная с концепцией сетей следующего поколения (Next Generation Networks, NGN) и внедрения платформы IMS (IP Multimedia Subsystem), оказывает существенное влияние на развитие СЭСН, функционирующих в интересах органов: государственной власти, обороны страны и безопасности государства.

В СЭСН активно внедряются современные услуги: видеоконференцсвязь, IP-телефония, электронная почта, передача файлов, веб-сервисы и другие, следовательно, становится задача обеспечения качества обслуживания пользователей – Quality of Service (QoS) [1].

Предоставление современных инфокоммуникационных услуг в СЭСН предполагает широкое использования IP-шифраторов и криптомаршрутизаторов совместно с технологией виртуальных частных сетей (Virtual Private Network, VPN).

Использование IP-технологий делает СЭСН во многом схожими с сетью электросвязи общего пользования (СЭОП), однако можно выделить отличительную особенность: в сети доступа в СЭСН есть высокоскоростные участки в десятки и сотни Мбит/с, а на транспортном уровне пропускная способность может существенно снижаться (до пропускной способности эквивалентной цифровому потоку Е1 – 2048 кбит/с), что создает эффект «бутылочного горлышка», тем самым ухудшая показатели качества обслуживания пользователей. Следовательно, необходимо применять дополнительные методы, учитывающие наличие низкоскоростных участков в СЭСН и позволяющие максимально и сбалансированно использовать весь ресурс сети, а не только ресурс отдельных каналов связи.

В теории построения инфокоммуникационных сетей данный класс методов получил название Traffic Engineering (TE – инжиниринг трафика). На сегодняшний день не существует универсального и стандартизированного подхода применения данных методов не только в СЭСН, но и в СЭОП.

Методы TE позволяют не только определить оптимальный маршрут для потока трафика, но и резервируют для него пропускную способность ресурсов сети, находящихся в этом маршруте [2].

На основе созданной аналитической модели VPN в СЭСН проведен сравнительный анализ используемых в СЭСН протоколов маршрутизации (RIP, OSPF) и перспективной для СЭСН технологии TE.

Сравнительный анализ позволил показать, что использование традиционных протоколов маршрутизации (RIP, OSPF) позволяет использовать до 40 % ресурса СЭСН, что обусловлено выбором единственного маршрута. Следствием одномаршрутности протоколов RIP и OSPF

является непропорциональная загрузка маршрутизаторов. На имитационной модели СЭСН [3] был провиден эксперимент (см. рис. 1), который показал, что наиболее загруженными являются маршрутизаторы 2, 4, 5, 8, 11, 12, 20, 21, 22, 23, что обусловлено топологическими особенностями СЭСН и скоростными параметрами используемых в СЭСН цифровых систем передачи. Применение в тех же условиях ТЕ позволяет добиться равномерного распределения нагрузки на сеть (см. рис. 1).

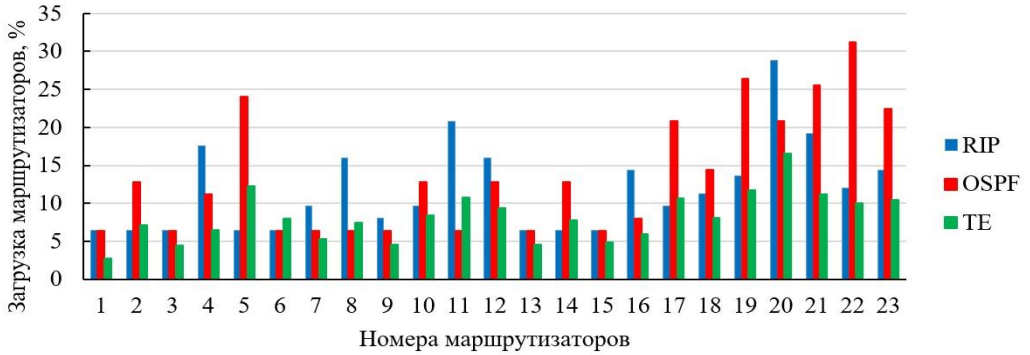


Рис. 1. Загрузка маршрутизаторов СЭСН

Таким образом, разработка научно обоснованных методов инжиниринга трафика для СЭСН при планировании VPN-туннелей, и их реализация совместно с современными концепциями QoS является сложной и актуальной научной задачей.

Потоковая модель VPN без учета задержки передачи пакета

Для реализации механизмов ТЕ в СЭСН, как и в СЭОП основным параметром является пропускная способность каналов связи между коммутационными устройствами – ресурс сети.

Определение ресурса сети для планирования VPN-туннелей в транспортной сети СЭОП осуществляется посредством применения потоковой модели VPN [4], где в качестве ресурса рассматривается пропускная способность,

$$\Delta = \boldsymbol{Q}_{\text{опт}} - \Lambda = \left[q_{ij_{\text{опт}}} - \lambda_{ij} \right], \forall i, j, \qquad (1)$$

где $q_{ij_{\text{опт}}}$ – минимальная пропускная способность оптимального маршрута $p_{\text{опт}}^{\langle ij \rangle}$ между маршрутизаторами i и j; λ_{ij} – интенсивность трафика между маршрутизаторами i и j.

Согласно выражению (1), ресурс сети по пропускной способности равен разности матрицы минимальных пропускных способностей для каждого оптимального пути $\boldsymbol{Q}_{\text{опт}}$ и матрицы интенсивности трафика между маршрутизаторами Λ.

Потоковая модель VPN с учетом задержки передачи пакета

Однако, с учетом требований по задержки передачи пакета, выражение (1) принимает вид:

$$\Delta^{\langle w_k \rangle} = \begin{cases} \Delta = \boldsymbol{Q}_{\text{опт}} - \Lambda\big|_{w^{\langle ij \rangle} < w_k \ \forall \ k=\overline{1,7}}, \\ w^{\langle ij \rangle} = w_{\text{во}}^{\langle ij \rangle} + w_{\text{р}}^{\langle ij \rangle} + w_{\text{обр}}^{\langle ij \rangle} + w_{\text{о}}^{\langle ij \rangle} + w_{\text{с}}^{\langle ij \rangle} \end{cases},$$

где W_k – пороговое значение задержки передачи пакета для определенного класса трафика $k = \overline{1,7}$, определенного в соответствии с [5]; $w^{\langle ij \rangle}$ – суммарная задержка передачи пакета между маршрутизаторами i и j; $w_{\text{во}}^{\langle ij \rangle}$ – задержка внеузловой обработки, которая задается в зависимости от кодека работы оконечного устройства; $w_{\text{р}}^{\langle ij \rangle}$ – задержка распространения пакета;

$w_{\text{обр}}^{\langle ij \rangle}$ – задержка обработки пакета маршрутизатором; $w_{\text{о}}^{\langle ij \rangle}$ – задержка ожидания пакета в буфере маршрутизатора; $w_{\text{с}}^{\langle ij \rangle}$ – задержка сериализации пакета.

Согласно предложенной модели запас по пропускной способности равен разности матриц $\boldsymbol{Q}_{\text{опт}}$ и $\boldsymbol{\Lambda}$, если задержка передачи пакета $w^{\langle ij \rangle}$ не достигнет порогового значения w_k, это позволяет не учитывать «несуществующий» ресурс сети по пропускной способности, ввиду невыполнения требований качества обслуживания. В свою очередь, используемая в настоящее время потоковая модель (1) показывает, что ресурс сети по пропускной способности есть (на рис. 2 представлен закрашенной областью) и может быть использован для планируемых VPN-туннелей.

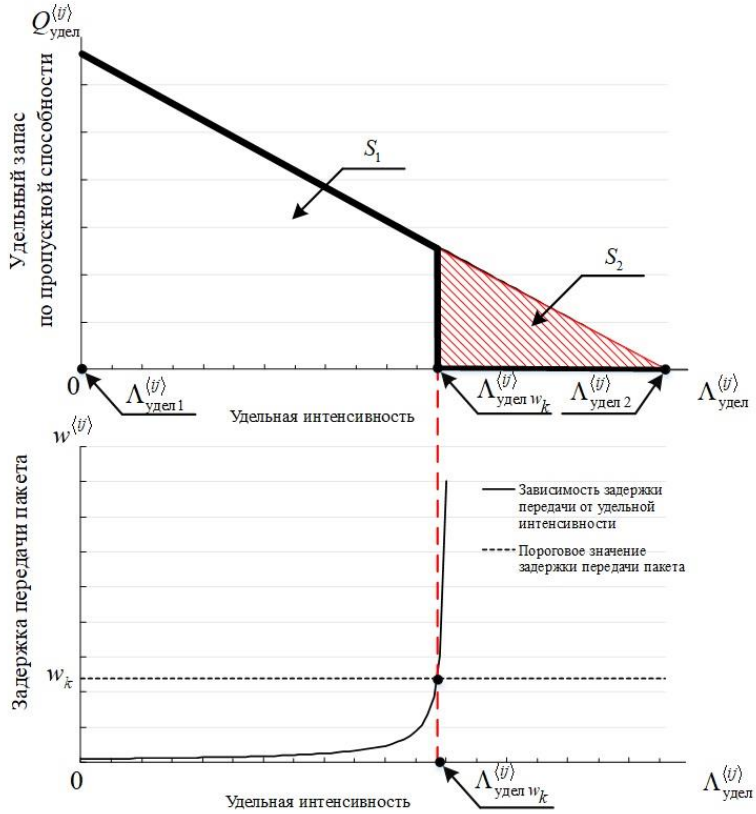


Рис. 2. Зависимости удельного запаса по пропускной способности и задержки передачи пакета от удельной интенсивности

На рис. 2 представлены следующие зависимости, расположенные соосно: зависимость удельного запаса по пропускной способности $Q_{\text{удел}}^{\langle ij \rangle}$ от удельной интенсивности трафика $\Lambda_{\text{удел}}^{\langle ij \rangle}$; зависимость задержки передачи пакета $w^{\langle ij \rangle}$ от удельной интенсивности трафика $\Lambda_{\text{удел}}^{\langle ij \rangle}$.

В СЭСН каждому организуемому VPN-туннелю предъявляются требования по пропускной способности и по задержки передачи пакета. Следовательно, использование модели (1) при планировании VPN-туннелей без учета задержки передачи пакета приводит к появлению ошибки определения запаса по пропускной способности для планируемых VPN-туннелей $k_{\text{ош}}$ (далее – ошибки).

Исходя из зависимости удельного запаса по пропускной способности от удельной интенсивности трафика, представленной на рис. 2, ошибка равна отношению

$$k_{\text{ош}} = \frac{S_w}{S_{\text{без}\,w}} \cdot 100\,\% = \frac{S_2}{S_1 + S_2} \cdot 100\,\%,$$

где $S_1 = \int_{\Lambda_{\text{удел}\,w_k}^{\langle ij\rangle}}^{\Lambda_{\text{удел}1}^{\langle ij\rangle}} Q_{\text{удел}}^{\langle ij\rangle}(\Lambda_{\text{удел}}^{\langle ij\rangle})d\Lambda_{\text{удел}}^{\langle ij\rangle}$ – площадь всей фигуры под кривой $Q_{\text{удел}}^{\langle ij\rangle}(\Lambda_{\text{удел}}^{\langle ij\rangle})$ до достижения допустимого порогового значения задержки передачи на отрезке $(\Lambda_{\text{удел}\,1}^{\langle ij\rangle}, \Lambda_{\text{удел}\,w_k}^{\langle ij\rangle})$; $S_2 = \int_{\Lambda_{\text{удел}\,w_k}^{\langle ij\rangle}}^{\Lambda_{\text{удел}2}^{\langle ij\rangle}} Q_{\text{удел}}^{\langle ij\rangle}(\Lambda_{\text{удел}}^{\langle ij\rangle})d\Lambda_{\text{удел}}^{\langle ij\rangle}$ – площадь фигуры под кривой $Q_{\text{удел}}^{\langle ij\rangle}(\Lambda_{\text{удел}}^{\langle ij\rangle})$ после достижения допустимого порогового значения задержки передачи на отрезке $(\Lambda_{\text{удел}\,w_k}^{\langle ij\rangle}, \Lambda_{\text{удел}\,2}^{\langle ij\rangle})$.

Для предлагаемой модели проведено исследование влияния на ошибку следующих факторов:

– изменение пропускной способности каналов связи между маршрутизаторами (рис. 3);

– изменение допустимой задержки передачи пакета для определенного класса трафика (рис. 4);

– изменение средней длины передаваемых IP-пакетов (рис. 5).

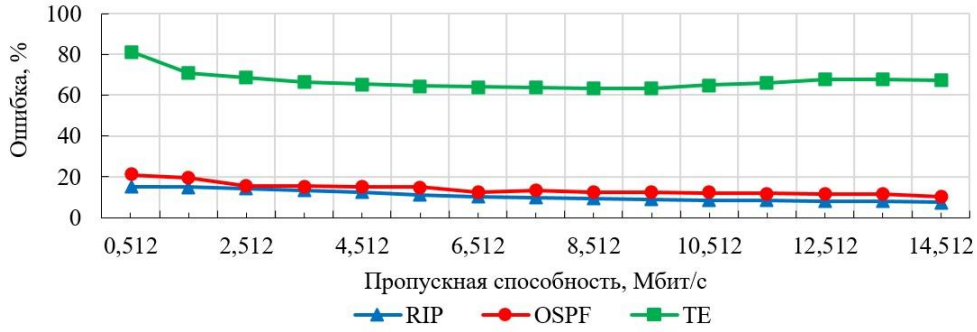


Рис. 3. Зависимость ошибки от пропускной способности каналов связи между маршрутизаторами

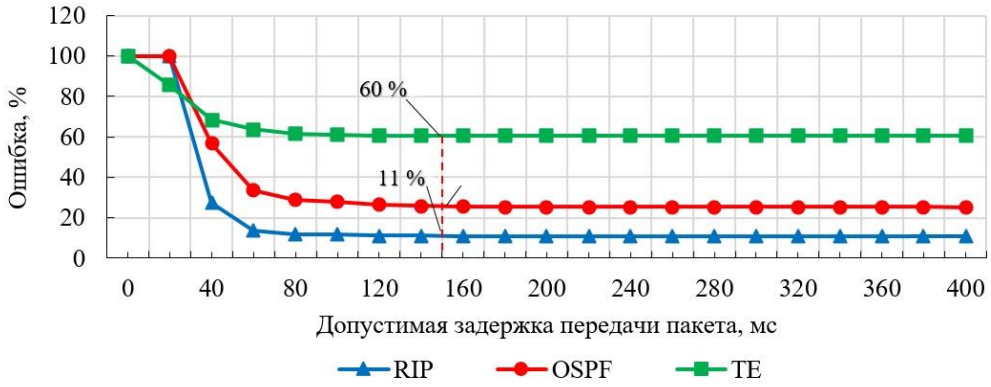


Рис. 4. Зависимость ошибки от допустимой задержки передачи пакета

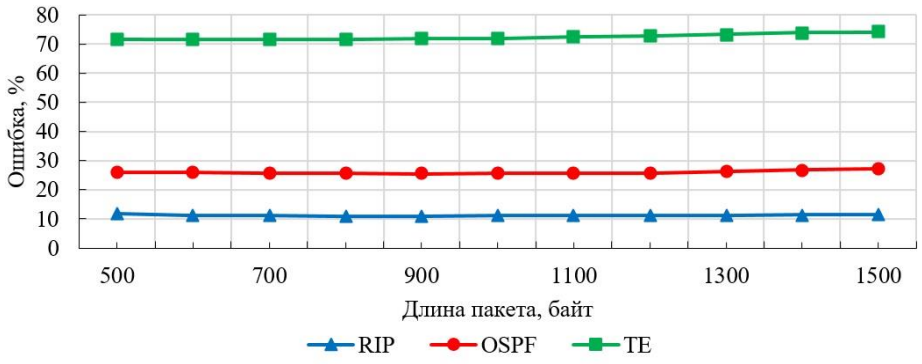


Рис. 5. Зависимость ошибки от средней длины IP-пакета

Анализ зависимостей, представленных на рис. 3, показывает, что при значительном увеличении пропускных способностей каналов связи между маршрутизаторами значение ошибки может быть равно нулю, что приведет к равенству величины запаса по пропускной

способности без учета задержки передачи пакета $\mathbf{\Delta}$ и запаса по пропускной способности с учетом задержки передачи пакета $\mathbf{\Delta}^{\langle w_k \rangle}$ – потоковая модель с учетом задержки вырождается в классическую потоковую модель VPN.

Диапазон изменения допустимой задержки передачи пакета от 0 – до 400 мс соответствует требованиям к QoS, представленным в [5] для интерактивного трафика. Из анализа зависимостей следует, что изменение допустимой значений задержки передачи пакета в рассматриваемом диапазоне не вносит значительных изменений в ошибку, что можно объяснить экспоненциальной зависимостью задержки передачи пакета от удельной интенсивности трафика (см. рис. 4) вследствие лавинообразного возрастания задержки ожидания пакетов в буфере маршрутизаторов при перегрузках в СЭСН. Для номинального значения допустимой задержки передачи пакета (150 мс), соответствующего нулевому классу трафика, ошибка при однопутевом маршрутировании (RIP, OSPF) составляет от 11 до 25 %, а при многопутевом – 60 %. Таким образом, для основного протокола маршрутизации (OSPF), используемого в СЭСН, существующая потоковая модель VPN вносит до 25 % ошибки при планировании VPN-туннелей, а для перспективной технологии TE ошибка составляет 60 %, что приводит к нерациональной оценке ресурса сети на этапе ее планирования.

Зависимости, представленные на рис.5, получены, при изменении длины IP-пакета от 500 до 1500 байт, что соответствует минимальной и максимальной длине пакета. Анализ зависимостей показывает, что увеличение длины пакета несущественно влияет на ошибку (изменение ошибки составляет от 2 до 4 %).

Заключение

Применение методов TE в СЭСН при планировании VPN-туннелей позволит обеспечить сбалансированную загрузку сети (до 100 % ресурса сети) в сравнении с традиционными одномаршрутными протоколами RIP и OSPF (обеспечивают использование до 40 % ресурса сети).

При необходимости передачи в планируемом VPN-туннеле интерактивного трафика возникает задача определения запаса по пропускной способности с учетом требований обеспечения качества обслуживания пользователей СЭСН. При этом, использование потоковой модели VPN приводит к существенной ошибке определения запаса по пропускной способности (при номинальном пороговом значении задержки 150 мс для одномаршрутных протоколов ошибка составляет от 11 до 25%, а для многомаршрутных – 60 %).

HOSE MODEL OF VPN WITH DELAY IN A COMMUNICATION NETWORKS OF SPECIAL PURPOSE

S.S. VRUBLEVSKY, A.A. BYSOV

Abstract. The article presents a hose model of VPN that takes into delay when planning VPN tunnels in communication networks of special purpose.

Keywords: communication networks of special purpose, hose model of VPN, delay, quality of service.

Список литературы

1. Абазина Е.С., Бусыгин А.В., Одоевский С.М. // Труды воен.-косм. акад. им. А. Ф. Можайского. 2020. № 672. С. 41–47.
2. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы. СПб. Питер, 2015.
3. Машкин Е.В., Бысов А.А., Врублевский С.С. // Весн. сувязі. 2022. № 5 (175). С. 68–72.
4. Росляков А.В. Разработка моделей и методов анализа виртуальных частных сетей с учетом особенностей их практической реализации : автореф. дис. д-ра техн. наук. Самара, 2008.
5. Требования к сетевым показателям качества для служб, основанных на протоколе IP. Женева, 2006.

UDC 004.81

HUMAN PHYSICAL ACTIVITY RECOGNITION ALGORITHM BASED ON SMARTPHONE DATA CONVOLUTIONAL NERUAL NETWOEK AND LONG SHORT TIME MEMORY

Z.X. YANG, Z.Y. CHEN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 19, 2023

Abstract. A deep learning framework for activity recognition based on smartphone acceleration sensor data, convolutional neural network (CNN) and long short-term memory (LSTM) is proposed in the paper. The proposed framework aims to improve the accuracy of human activity recognition (HAR) by combining the strengths of CNN and LSTM. The CNN is used to extract features from the acceleration data and the LSTM is used to model the temporal dependencies of the data. The proposed framework is evaluated on the publicly available dataset, it includes 6 different actions: walking, walking upstairs, walking downstairs, sitting, standing, and laying. The recognition accuracy has reached 94 %.

Keywords: HAR, CNN, LSTM, acceleration sensor.

Introduction

With the rapid popularity of smartphones and the rapid development of micro sensors, various MEMS sensor devices are embedded in people's smartphone. The main advantages of MEMS sensors are small size, light weight, low power consumption, high reliability, high sensitivity, easy integration, etc. In addition, the research on human behavior recognition based on sensor-based intelligent devices has become an emerging research topic in recent years, traditional machine learning algorithms extract feature vectors from data to distinguish between classes of activities, and researchers have done a lot of research work in this area. FGD Silva [1] used two methods, FDR and PCA, to extract 19 features from the data and used SVM method to classify the activities, VNT Sang [2] applied KNN, ANN and SVM algorithms for classification and recognition of human behavior, R Singh [3] et al. proposed an algorithm for human state recognition activity using decision tree C4,5 by data mining algorithm. Since traditional machine learning algorithms require manual extraction of features in the data, and it is difficult for non-professionals to extract effective feature sets, manual extraction is also subject to human error and time-consuming, all of these methods which will reduce the accuracy of classification and recognition, but neural networks greatly compensate for the lack of manual feature extraction in traditional machine learning by building a multi-level automatic feature extraction architecture.

Data acquisition for human behavior recognition is basically divided into two categories, video image-based data acquisition and wearable sensor-based data acquisition. In this paper, we focus on human activity recognition based on wearable sensing data. This paper uses the built-in acceleration sensor of a smartphone for data collection.

The network model consists of three convolutional layer and LSTM network layers, fully connected layer, and one Softmax layer, and predicts the corresponding human actions from the data set. The proposed algorithm's accuracy achieved 94 % and the loss is about 0,02 at the test set.

Dataset

We selected the data set from the UCI Human Activity Recognition Using Smartphones Data Set. The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities wearing a smartphone on the waist. Using its

embedded accelerometer, they captured 3-axial linear acceleration at a constant rate of 50 Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets.

The sensor signals (accelerometer) were pre-processed by applying noise filters and then sampled in fixed width sliding windows of 2,56 sec and 50 % overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0,3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

After the data set is obtained, the maximum and minimum values of the data are normalized using Formula (1). The data in the training set is 7352×561, and 7350×560 data are selected as the data set required for the experiment. It was found that many data with the same output in the data set were connected, which did not meet the assumption that the training data were independent and identically distributed. randon.seed = 314 was selected to generate a random sequence of length 7350, which was used as the index of the data:

$$X_{scale} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

where $X_{\max}$ – denotes the maximum value in the feature, $X_{\min}$ – denotes the minimum value in the feature.

Framework of CNN and LSTM algorithm

The structure diagram of the algorithm is shown in Figure 1. Based on the cell phone tri-axis acceleration data collected from the public dataset, the algorithm predicts six different human behaviors through convolutional layer and LSTM, fully connected layer, and Softmax layer, and finally outputs the behavior type with the highest probability as the output result.

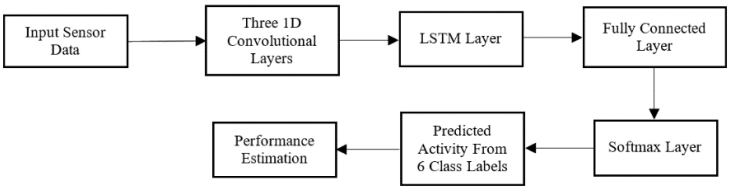


Figure 1. The network structure of CNN-LSTM

1D convolutional layer 1

For the convolutional 1, we got the input tensor is (100,1,560), the batch size is 100, and input channel is 1, length of signal sequence is 560, the hyperparameter of the convolutional layer is shown as Table 1.

In the 1D convolutional layer, it has two kinds of parameters: weights and biases. The total number of parameters is just the sum of all weights and biases:

$$W_c = K \times C \times N, \tag{2}$$

where the W_c – is the number of weights of the convolutional layer, C – is the number of channels of the input, N – is the number of kernels, K – is the size of kernels used in the convolutional layer, so we can get the weight of the convolutional layer is 320:

$$P_c = W_c + B_c, \tag{3}$$

where P_c – is the number of parameters of the convolutional layer, B_c – is the number of biases of the convolutional layer, W_c – is the number of weights of the convolutional layer, so we can get the number of parameters of the convolution layer 1 is 384.

The calculation method of size (O) of the output is shown in formula (4):

$$O = \frac{I - K + 2P}{S} + 1, \tag{4}$$

where the I – is the size (width) of input, K – is the size (width) of kernels used in the convolutional layer, S – is the stride of the convolution operation, P – is padding value. Moreover, we can get the size of the output is 278.

Table 1. **The hyperparameter of the 1D convolution layer 1**

Hyperparameter	Value
Input channel	1
Output channel	64
Kernel size	5
Stride	2
Padding	0

1D convolutional layer 2

For the convolutional 2, we got the input tensor is (100,64,278), the batch size is 100, and input channel is 64, length of signal sequence is 278, the size of the output is 137. The parameter and hyperparameter of the convolutional layer are shown as Table 2 and Table 3.

Table 2. **The parameter value of the 1D convolution layer 2**

Parameter	Value
Weight	20480
Bias	64
Total	20544

Table 3. **The hyperparameter of the 1D convolution layer 2**

Hyperparameter	Value
Input channel	64
Output channel	64
Kernel size	5
Stride	2
Padding	0

1D convolutional layer 3

For the convolutional 3, we got the input tensor is (100,64,137), the batch size is 100, and input channel is 64, length of signal sequence is 137, the size of the output is 67. The parameter and hyperparameter of the convolutional layer are shown as Table 4 and Table 5.

Table 4. **The parameter value of the 1D convolution layer 3**

Parameter	Value
Weight	20480
Bias	64
Total	20544

Table 5. **The hyperparameter of the 1D convolution layer 3**

Hyperparameter	Value
Input channel	64
Output channel	64
Kernel size	5
Stride	2
Padding	0

LSTM layer

In the LSTM layer, input size is taken as 64 because the feature used is the 3rd convolutional output feature, which is the dimension 64, using convolutional features for temporal recognition, and the number of hidden unit size is 128. For the input tensor shape (100,64), after processing through the LSTM layer, the output shape become (100,128). The parameters are shown in Table 6. The hyperparameter of the layer are shown in Table 7, each LSTM has 128 hidden units, and that is, the number of neurons in the LSTM unit is 128.

Table 6. **The parameter value of the LSTM layer**

Parameter	Value
Input tensor shape	(100,64)
LSTM units	128
Output tensor shape	(100,128)

Table 7. **The hyperparameter of the LSTM layer**

Hyperparameter	Value
Hidden layer	3
Hidden unit size	128
Learning rate	0.001
Dropout	0.9
Batch size	100
Epoch	50

Fully connected layer

In the fully connected layer, we get the input tensor derived from the output of the last hidden layer state of the LSTM layer, the input shape is (100,128).

After we got the input, we can multiply the input by the weight matrix [128,6] and add a bias [6], then we get the output of the fully connected layer (100,6), the relevant parameters are shown in Table 8 and the calculation process is shown in formula (5):

$$y = xA + b, \tag{5}$$

where x – is the input, A – is the weight matrix, b – is the bias.

Table 8. **The input, condition, and output of fully connected layer**

Input shape	(100,128)
Weight matrix	[128,6]
Bias	[6]
Output shape	(100,6)

Softmax layer

After getting the output from the fully connected layer, we use the activation function Softmax to map the output of the neuron to the (0,1) interval, which allowed us to perform multiple classification:

$$\text{Soft}\max\left(Z_i\right) = \frac{e^{Z_i}}{\sum_{c=1}^{C} e^{Z_c}}, \tag{3}$$

where Z_i – is the output value of the i class, C – is the number of output nodes, which also represents the number of categories in the classification. In this dataset, because there are 6 feature actions in this dataset, we set the C value to 6.

Evaluation indicators and results

Evaluation refers to the stage where the results of model testing are measured and evaluated. We measure the performance of the model using a confusion matrix, which is also used to determine accuracy. The confusion matrix which we used is shown in Table 9.

Table 9. **Confusion matrix of the six classifications tasks**

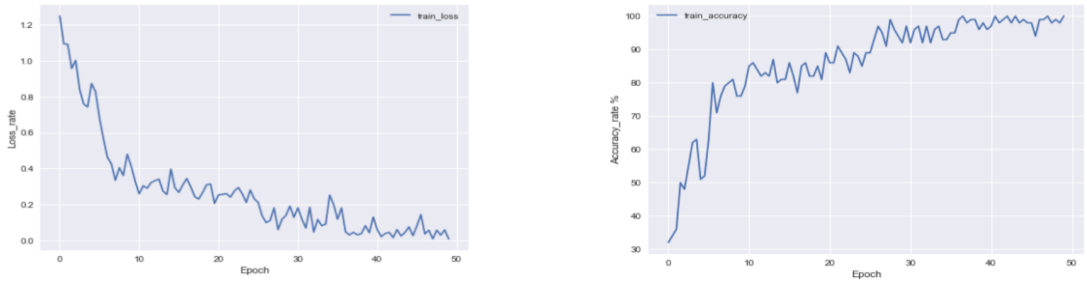
Six-class label		Predicted label						Total
		Laying	Walking	Upstairs	Downstairs	Sitting	Standing	
True label	Laying	A_{LL}	A_{LW}	A_{LU}	A_{LD}	A_{LST}	A_{LSD}	T_L
	Walking	A_{WL}	A_{WW}	A_{WU}	A_{WD}	A_{WST}	A_{WSD}	T_W
	Upstairs	A_{UL}	A_{UW}	A_{UU}	A_{UD}	A_{UST}	A_{USD}	T_U
	Downstairs	A_{DL}	A_{DW}	A_{DU}	A_{DD}	A_{DST}	A_{DSD}	T_D
	Sitting	A_{STL}	A_{STW}	A_{STU}	A_{STD}	A_{STST}	A_{STSD}	T_{ST}
	Standing	A_{SDL}	A_{SDW}	A_{SDU}	A_{SDD}	A_{SDST}	A_{SDSD}	T_{SD}

Each sample in the classification task has only one defined category, and a prediction of that category is a correct classification, and a failure to predict it is a classification error, so the most intuitive metric is Accuracy. The formula is as follows:

$$Accurancy = \frac{A_{LL} + A_{WW} + A_{UU} + A_{STST} + A_{SDSD}}{T_L + T_W + T_U + T_D + T_{ST} + T_{SD}}, \quad (4)$$

where the A_{LL} – denotes the accuracy of correct prediction of laying, A_{WW} – denotes the accuracy of correct prediction of walking, A_{UU} – denotes the accuracy of correct prediction of upstairs, A_{DD} – denotes the accuracy of correct prediction of downstairs, A_{STST} denotes the accuracy of correct prediction of sitting, A_{SDSD} – denotes the accuracy of correct prediction of laying.

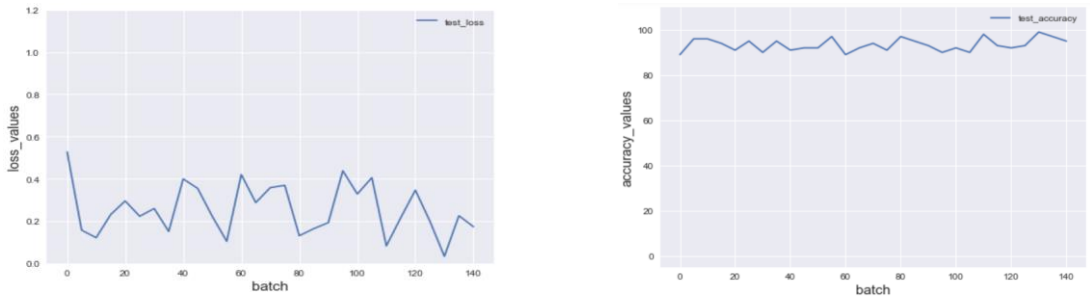
Our model is training and testing on the UCI HAR dataset, and 70 % of the volunteers was selected for generating the training data and the remaining 30 % is used to test the data. We set the training set to 100 and train on the training set for 50 epochs. The loss value and the accuracy values are shown in Figure 2.



a *b*

Figure 2. Loss and accuracy rate of the CNN-LSTM model: *a* – loss rate during the training; *b* – accuracy rate during the training

Next, we will apply the trained model to the test set to see how well the model performs on the test set. Figure 3 visualizes the accuracy and loss rate of the model on the test set.



a *b*

Figure 3. Loss and accuracy rate of the CNN-LSTM model: *a* – loss rate during the testing; *b* – accuracy rate during the testing

It can be seen from the test set that there is some fluctuation in the loss rate of the model, since the batch size is only 100, it does not look smooth on the test set, but it still has smaller loss rate and higher accuracy on average. Our proposed model achieves an average recognition accuracy of 94 % on the training set, while the average value of loss is about 0,25.

Table 10. **Confusion matrix of the six-class dataset scheme on the test dataset**

Six-class label		Predicted label					
		Laying	Walking	Upstairs	Downstairs	Sitting	Standing
True label	Laying	1	0	0	0	0	0
	Walking	0	0,970	0,0061	0,022	0	0
	Upstairs	0	0,067	0,900	0,037	0	0
	Downstairs	0	0,044	0,093	0,860	0	0
	Sitting	0	0	0	0	0,940	0
	Standing	0	0	0	0	0,086	0,910

The confusion matrix in Table 10 is used to determine the performance of the trained model on the six human action categories. It displayed the accuracy of each class where laying is 100 %, Walking is 97 %, Walking upstairs is 90 %, Walking downstairs is 86 %, sitting is 94 %, and standing is 91 %. Next, normalize the confusion matrix in Figure 4.

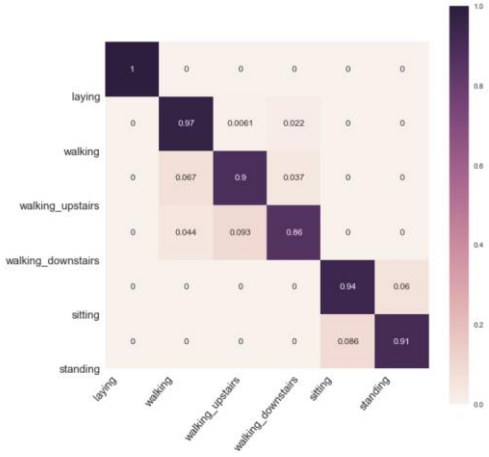


Figure 4. Normalized confusion matrix of six class dataset schemes

Conclusion

This paper is based on the UCI HAR dataset, the dataset was collected by using a smartphone placed around the waist of the tested volunteers, it includes six different human behavioral states: walking, walking upstairs, walking downstairs, sitting, standing, and laying. The proposed model includes three 1D convolutional layers, LSTM network layer, fully connected layer, and Softmax layer. The algorithm extracts data features from the input signal sequence using the three-layer convolutional neural network, and then uses the features as the input of the LSTM neural network. After obtaining the final output of the LSTM neural network, it is mapped to the fully connected layer. Finally, the output is transformed into the probability corresponding to the state through the Softmax layer. With the training of this neural network, our algorithm achieves an average accuracy of 94 % for the six feature activities.

References

1. Silva F.G., Galeazzo E. // Accelerometer based intelligent system for human movement recognition. 2013. P. 20–24.
2. Vnt Sang., Vu Ngoc Thanh. // Human activity recognition and monitoring using smartphones. 2015. P. 481–485.
3. Singh R., Kumar H., Singla R.K. // Analysis of Feature Selection Techniques for Network Traffic Dataset. 2014. P. 42–46.

УДК 519.724

АЛГОРИТМ КОДИРОВАНИЯ ПРОЦЕССА ТРАНСЛЯЦИИ БЕЛКОВ В КЛЕТКЕ

М.А. ПРОТЬКО, О.Ф. БОРИСЕНКО

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 19 марта 2023

Аннотация. Целью данной статьи является рассмотрение азотистых оснований в соотношении с кодируемыми ими аминокислотами, с дальнейшими поисками их взаимосвязи.

Ключевые слова: четверичный код, генетический код, трансляция.

Введение

Если поставить целью создание динамически развивающейся системы (системы, способной реагировать на условия, изначально не предусмотренные при ее проектировании, но потенциально возможные [1]), необходимо четко разграничить все ее параметры, или же, поставить строго структурированную формальную задачу.

Для достижения поставленной цели рассмотрим структуру генетического кода, свойственного ДНК и РНК, для выявления неких закономерностей.

В данной работе используются определения как из теории кодирования, так и из общей генетики. В начале следует определение из теории кодирования, затем, в скобках, из общей генетики.

Определение объекта

Рассмотрим предметную область. Генетическим кодом называется последовательность четырех азотистых оснований (аденин (А), тимин (Т) / урацил (У), гуанин (Г) и цитозин (Ц)) которым в соответствие ставиться 20 аминокислот (см. табл. 1).

Определим генетический код следующим образом, воспользовавшись [2]:

Положим, что существует некий источник, выдающий дискретное сообщение a (полипептиды и/или белки), которое можно рассматривать как последовательность элементарных сообщений a_i (аминокислоты). Эти элементарные сообщения – символы, и их совокупность $\{a_i\}$ – алфавит.

Пусть последовательность символов источника a заменяется последовательностью кодовых символов (триплетом, или же кодоном).

Элементарными символами кодовой комбинации в данном случае служат азотистые основания.

Общее число символов, составляющих кодовую комбинацию (длина кода) $n = 3,$ количество значений кодовых признаков (основание кода) $m = 4.$

Емкость кода $N_o = 64.$

Количество сообщений $N_a = 20.$

Кодовое расстояние $d_0 = 1.$

Относительная скорость кода $R_k = \log_2 N_a / \log_2 N_0 = 2,996 / 4,159 = 0,720.$

Избыточность $c_k = 1 - R_k = 0,280.$

Табл. 1. **Соответствия аминокислота – триплет**

Название	Частота	Кол-во кодонов	Кодоны	Мин. к.р.
Лейцин Leu	9,68	6	УУА УУГ ЦУУ ЦУЦ ЦУА ЦУГ	2 2
Аланин Ala	8,76	4	ГЦУ ГЦЦ ГЦА ГЦГ	1–2
Серин Ser	7,14	6	УЦУ УЦЦ УЦА УЦГ АГУ АГЦ	2 1
Глицин Gly	7,03	4	ГГУ ГГЦ ГГА ГГГ	1
Валин Val	6,73	4	ГУУ ГУЦ ГУА ГУГ	1
Глютаминовая кислота Glu	6,32	2	ГАА ГАГ	2
Аргинин Arg	5,78	6	ЦГУ ЦГЦ ЦГА ЦГГ АГА АГГ	2 1
Треонин Thr	5,53	4	АЦУ АЦЦ АЦА АЦГ	2
Асппарагиновая кислота Asp	5,49	2	ГАУ ГАЦ	2
Изолейцин Ile	5,49	3	АУУ АУЦ АУА	2
Лизин Lys	5,19	2	ААА ААГ	2
Пролин Pro	5,02	4	ЦЦУ ЦЦЦ ЦЦА ЦЦГ	2
Аспарагин Asn	3,93	2	ААУ ААЦ	2
Глютамин Gln	3,90	2	ЦАА ЦАГ	3
Фенилаланин Phe	3,87	2	УУУ УУЦ	1
Тирозин Tyr	2,91	2	УАУ УАЦ	3
Метионин Met	2,32	1	АУГ	3
Гистидин His	2,26	2	ЦАУ ЦАЦ	2
Цистеин: Cys	1,38	2	УГУ УГЦ	1
Триптофан Trp	1,25	1	УГГ	–

Генетический код (далее г.к.) является равномерным ($n = const$) и многопозиционным (хромосомный г.к.), однако, существует и неравномерный г.к. (митохондриальный г.к.).

По форме представления в канале передачи (процесс кодирования, или же трансляции, переход РНК – белок) – г.к. имеет параллельную форму.

По основным законам кодообразования, г.к. – комбинаторный код.

Рассмотрим табл. 1, являющуюся объединением информации из источников [3] и [4].

В табл. 1 в столбце «название» находится аминокислота с ее русским названием и английским сокращением, частота встречаемости основана на выборке из 7 555 843 062 аминокислот. Данные частоты являются усредненными показателями по всем царствам.

Положим за кодовое расстояние количество различающихся элементарных символов в кодоне.

Минимальное кодовое расстояние (Мин. к.р.) – разница между элементарными символами пары кодонов, рассматриваемых по их частоте встречаемости (Leu – Ala, Ala – Ser и т.д.). В случае, если аминокислота имеет кодоны, кодовое расстояние между которыми более 1, данные кодоны разделяются на группы. К каждой такой группе записано свое кодовое расстояние, рассчитанное по аналогичному принципу.

Как видно из табл. 1, количество кодовых групп (кодонов) никак не зависит от частоты встречаемости, кодируемой ими аминокислоты.

Кодовое расстояние между кодонами, кодирующими одно и то же основание, чаще всего, не превышает 1. Исключения: аргинин, серин и лейцин (6 кодовых последовательностей).

Кодовое расстояние между разными основаниями чаще всего – 2.

Никакой ярко выраженной закономерности между частотой встречаемости и кодовым расстоянием не наблюдается.

Из вышеописанного можно сделать вывод, что г.к. не является оптимальным (одно из свойств вырожденности).

Г.к. является помехоустойчивым кодом, поскольку позволяет при обнаружении ошибочной последовательности завершать трансляцию (таких последовательностей всего три, это «стоп» кодоны).

Г.к. очень близок к полному коду, согласно определению из [5].

Если положить, что «бессмысленные» последовательности («стоп» кодоны) являются разрешенными, то г.к. – полный код.

Рассмотрим пример части последовательности 11 хромосомы человека [6]:

АУГ ГУЦ ЦУГ ГГУ ГГЦ АУГ ГАГ ЦУЦ УУГ ЦАЦ ЦУЦ УАГ Г…

Met Val Leu Gly Gly Tyr Glu Leu Leu His Leu Стоп

Если предположить, что такая последовательность действительно кодирует некий белок, можно сделать следующий вывод: один и тот же кодон в случае, если он относится к группе избыточных, не будет повторяться.

Рассмотрим последовательность, преобразованную циклическим сдвигом:

ЦУУ ГЦУ ГГУ ГАА ЦГУ

Leu Ala Gly Glu Arg

УУГ ЦУГ ГУГ ААЦ ГУЦ

Leu Leu Val Thr Val

Можно заметить, что полученная таким образом комбинация все еще имеет смысл.

Если определить операции над множеством элементарных сигналов, а также матрицу строк, являющуюся аналогией матрицы полного кода, с линейной независимостью строк, можно сказать, что г.к. – циклический код.

Если же определить операции над множеством элементарных сигналов согласно свойствам, описанным в [7] и с учетом описанных алгоритмов в [6], получим, что г.к. – четверичный код.

При определении операции над множеством элементарных сигналов стоит отталкиваться от смысла изначального алфавита (аминокислоты и «текста» белков), поскольку от него будет зависеть образующая строка матрицы разрешенных последовательностей циклического кода.

Если предположить, что в нашей системе, где будет использоваться алгоритм г.к., существует процесс, подобный мутации и кроссинговеру, то количество запрещенных

последовательностей характеризует выраженность данного процесса. Чем меньше число «стоп» кодонов, тем больше избыточность, или же, тем меньше стабильность.

Пример на основании алгоритма Шенона-Фоне

Рассмотрим способ замены последовательности символов источника a (20 аминокислот) кодовыми символами с $n=3$ и $m=4$. Воспользуемся алгоритмом Шенона-Фоне. Положим, что частота встречаемости символов, как и сами символы источника, аналогичны представленным в табл. 1.

Результаты представлены в табл. 2.

Табл. 2. **Соответствия при *n*=3**

Название	Кодоны	Кол-во кодонов	Мин. к.р.
Лейцин Leu	ЦЦ-	4	1
Аланин Ala	ЦГ-	4	1
Серин Ser	ЦУ-	4	2
Глицин Gly	ГЦ-	4	1
Валин Val	ГГ-	4	1
Глютаминовая кислота Glu	ГУ-	4	1
Аргинин Arg	ГА-	4	2
Треонин Thr	УЦ-	4	1
Асппарагиновая кислота Asp	УГ-	4	1
Изолейцин Ile	УУ-	4	1
Лизин Lys	УАЦ	1	1
Пролин Pro	УАГ	1	2
Аспарагин Asn	АЦ-	4	1
Глютамин Gln	АГ-	4	1
Фенилаланин Phe	АУЦ	1	1
Тирозин Tyr	АУГ	1	2
Метионин Met	ААЦ	1	1
Гистидин His	ААГ	1	1
Цистеин: Cys	ААУ	1	1
Триптофан Trp	ААА	1	–

Получаем код из 56 значащих последовательностей и 8 запрещенных. Стоит учитывать то, что минимальное кодовое расстояние в данном случае у большинства символов источника – 1 (посчитанное по тому же принципу, что и в табл. 1).

Пример на основании кодового расстояния

Рассмотрим способ замены последовательности символов источника кодовыми символами с учетом кодового расстояния.

Пусть имеется последовательность символов источника a (Leu, Ala, Ser, Gly, Val).

Сохраняя соотношения, между емкостью кода и количеством символов источника, получим емкость кода $N_o = 16$.

Таким образом, длина кода $n=2$ с основанием $m=4$.

Полный код при данных условиях:

$$\pi_4^2 = \begin{vmatrix} А\,А\,А\,А\,Г\,Г\,Г\,Г\,У\,У\,У\,У\,Ц\,Ц\,Ц\,Ц \\ А\,Г\,У\,Ц\,А\,Г\,У\,Ц\,А\,Г\,У\,Ц\,А\,Г\,У\,Ц \end{vmatrix}.$$

Выбор последовательности кодовых символов из полного кода для символов источника будем совершать таким образом, чтобы кодовое расстояние между двумя соседствующими по частоте встречаемости символами источника было максимальным.

Таким образом, получим соотношения, представленные в табл. 3.

Табл. 3. **Соответствия при *n*=2**

Leu	Ala	Ser	Gly	Val
АЦ	ГЦ	АА	ГГ	УУ
АУ	ГА	АГ	ГУ	УГ
ЦЦ	УЦ	ЦА	ЦГ	УА

В табл. 3 представлен один из вариантов кода, получаемого из полной последовательности. В данном случае, самые отличающиеся по характеристикам варианты – те, у которых разное число запрещенных последовательностей, (число «стоп» кодонов – от 1 до 6).

Заключение

Ключевое свойство, приводящее одновременно и к стабильности, и к изменчивости систем на основе генетического кода – это его избыточность. Причем избыточность такого рода, что при возникновении ошибки (мутации) вероятность критических изменений смысла сообщения остается достаточно малой (кодон либо переходит в свой эквивалент, либо в иную существующую аминокислоту). Т.е., для построения системы с подобным свойством достаточно выбора подходящего веса, рассматривая ошибки как благо. Разрядность кода не играет столь существенной роли в достижении этого свойства.

Дальнейший анализ полученных кодовых последовательностей заключается в сборе статистики при использовании данных кодовых последовательностей в генетических алгоритмах, по таким параметрам как количество популяций на алгоритм кодирования (отсчет популяции заканчивается при нахождении решения или вырождении).

ALGORITHM FOR ENCODING THE PROCESS OF PROTEIN TRANSLATION IN A CELL

M.A. PROTSKO, O.F. BORISENKO

Abstract. The aim of this article is to find correlation between nitrogenous bases in relation to the amino acids they encode.

Keywords: quaternary code, genetic code, translation.

Список литературы

1. Протько М.А., Борисенко О.Ф. // Простейшие шифры и генетический алгоритм. 2023.
2. Кузьмин И.В., Кедрус Н.А. Основы теории информации и кодирования. 1986.
3. Каминская Э.А. Общая генетика. Минск, Вышэйшая школа. 1992.
4. Kozlowski L.P. // Proteome-pI: proteome isoelectric point database. Nucleic Acids Research. 45 (D1): D1112–D1116. doi:10.1093/nar/gkw978. PMC 5210655. PMID 27789699
5. Варакин Л.Е. Системы связи с шумоподобными сигналами. Москва, Радио и связь.1985.
6. Фрагмент генетичекого кода из 11 хромосомы человека [Электронный ресурс]. URL: http://www.ensembl.org/Homo_sapiens/Info/Index?db=core
7. Зиновьев Д.В., Соле П. // Пробл.передачи информ. 2004, Т. 40, В. 2. С. 50–62.

UDC 621.391

RESEARCH ON WIRELESS AD HOC NETWORK TECHNOLOGY

Y. WANG, P. ZENG, Z.J. WEI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 20, 2022

Abstract. With the rapid development of communication technology and the improvement of people's requirements for communication, wireless AD hoc network has become an important research content of today's network. Wireless AD hoc network (WLAN) is a multi-hop, flexible, non-center network formed by several mobile wireless nodes self-organization. These characteristics provide a favorable guarantee for the military and civilian communication field. This paper first introduces the characteristics and applications of wireless AD hoc network. Secondly, the problems of existing routing protocols are studied. Finally, according to the problems of energy consumption, multipath and multicast in wireless AD hoc network, six methods are summarized to optimize the system and improve the transmission performance of wireless AD hoc network.

Keywords: wireless AD hoc network; routing protocol; rerformance optimization.

Introduction

Wireless AD hoc network is a multi hop, no center, temporary system. Each node in the network can be configured for fast networking, short networking time, and low system cost. It can work independently or with other networks in the form of subnets. The management of AD hoc network is relatively simple, without the control of the central base station, so it has good robustness and flexibility. In the military field, civil aviation technology has been widely used and has become a feasible or even the only feasible communication solution for temporary situations. Its characteristics are as follows:

1. No center: Without the support of base station and other control centers, each node in the network is equal.
2. Self-organization: After a node is powered on, it can automatically discover neighbor nodes for fast networking.
3. Multi-hop: When two communication nodes are not within the transmission range, the intermediate node can be forwarded.
4. Dynamic topology: Nodes can be moved randomly. Nodes in the network can join or leave at any time, resulting in changes in the topology structure.
5. Limited energy: AD hoc networks are often used in temporary situations. Most nodes are powered by batteries, playing the dual identity of terminal nodes and relay nodes. The exhaustion of energy will not only make a single node fail, but also may change the whole network topology.
6. Security: Without the help of a trusted third party, it is vulnerable to link layer attacks, eavesdropping and damage.

According to the above characteristics, it can be found that the wireless AD hoc network has great advantages compared with other networks, but there are some shortcomings in network security and energy saving. In order to make the wireless AD hoc network more reliable and bring greater benefits to our life, routing protocol is the core and lifeblood of the network and efficient and reliable routing protocol is the guarantee of the normal operation of the system.

Protocols for wireless AD hoc networks

According to the route discovery policy, routing protocols are divided into three types: table-driven routing protocols, on-demand routing protocols, and hybrid routing protocols, as shown in Figure 1.

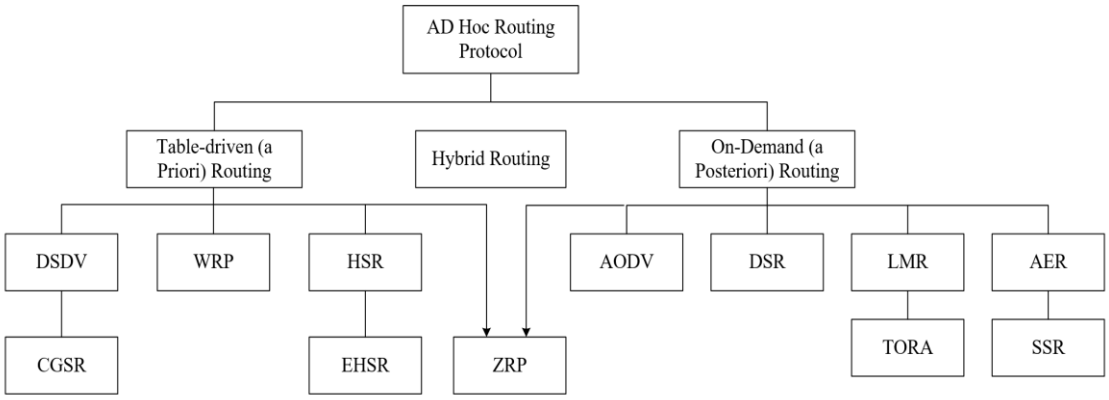


Figure 1. Common routing protocols for AD Hoc networks

On-demand Routing Protocol. This protocol includes route discovery and route maintenance. Nodes do not need to exchange information at any time, but maintain routing messages. When a node sends data packets, the route discovery function is enabled and the packets are forwarded in the form of flooding. Disadvantages: Latency of the initial route is a problem because the routing node looks up the route on demand without storing the route information beforehand. In some emergencies, the message delay may be delayed to the event because great influence. Advantages: Since there is no need to periodically broadcast, the network overhead is reduced.

Table Driven Routing. ProtocolEach node needs to maintain a routing table to transmit data according to the routing table, update the routing table in real time, and establish routes from the source node to the destination node. Each node in the network maintains a routing message table to the other nodes, so each node must have a high storage capacity. Disadvantages: Nodes need to master the topology of the whole network, so it is only suitable for small-scale networks. Nodes are always in working state, which will increase the network control cost and reduce the network life time. Advantages: The protocol always establishes routing links, and the delay between nodes is small. It can quickly adapt to network topology changes, and the packet loss rate to the system is small.

Hybrid Routing. Based on the analysis of the above performance and their respective advantages, a hybrid routing protocol is developed which is more excellent than each other. To improve the transmission efficiency of wireless AD hoc network system, the research results of literature [1] show that hybrid routing protocol (DSR+ FSR) has the characteristics of fast convergence, strong adaptability of network environment, stable and reliable routing information, and is more suitable for complex network environment.

Method of routing protocol optimization

Wireless AD hoc network is a mobile network. The topology changes constantly, which degrades the network performance. Therefore, it is crucial to select the appropriate routing protocol for different environments. The routing protocol is improved from the perspectives of service quality, energy saving, security, multipath, multicast, topology and so on. Aiming at these shortcomings in the system, six technologies are summarized to optimize wireless AD hoc networks, including packet aggregation technology, local repair, local spanning tree, network coding technology, cross-layer routing, and self-adaptation.

Packet aggregation technology, the technology is to reduce the number of nodes to send packets. Instead of immediately forwarding data every time, nodes wait for the maximum aggregation time. During the aggregation time, new data packets are added to the queue, waiting for the data to be sent. Especially in areas with dense nodes, this data transmission mode will reduce the probability of collision.

In addition, several data packets are aggregated into one UDP broadcast packet to reduce the transmission times of data. The new routing protocol BATMAN adv uses packet aggregation technology to reduce the system overhead and improve the system utilization.

In the local repair technology, when the link between nodes changes, it should inform other nodes of the change as soon as possible, readjust and calculate the shortest path. The faster the link status changes, the more overhead is incurred. The local repair technology can reduce the network delay. Local link repair technology pays more attention to route connectivity, reduces route cost and shorens route recovery time. Literature [2] proposes a new multi-metric wireless routing protocol based on AODV, which comprehensively considers four factors, including minimum hops, residual energy, and energy loss rate and network node density. Most importantly, a low-cost and efficient repair strategy is introduced to optimize AODV and improve the performance of routing protocols in network systems.

For broadcasting based on local spanning tree, routing protocols all adopt flooding technology to improve the efficiency of flooding and reduce the cost of flooding. The optimal method of flooding technology is minimum spanning tree for broadcasting, but minimum spanning tree must master the topology of the whole network, so it is not advisable before the establishment of routes. Therefore, it is possible to broadcast through local spanning tree, and broadcast in the form of packet, which can reduce the cost of the network and increase the service life of the network compared with the simple use of flooding. OLSR routing protocols flood broadcast link information using a local spanning tree approach.

Network coding technology, which can improve system throughput. Aiming at the broadcast characteristics of the physical layer of wireless AD hoc networking, scholars have proposed a network coding routing protocol [3-5]. The protocol changes the previous channel mode, fuses a large number of packets encoded by each target node, and then sends the packets to each target node, improving the throughput of the network. Peng Yongxiangzai [6] proposed a code-aware unicast routing protocol for multi-hop wireless networks. This protocol effectively describes network programming in a special way

The routing measures of code and unicast session characteristics, and the routing protocol is improved. In order to ensure that this routing metric can be effectively combined with widely used routing algorithms, a unique mapping procedure is used to ensure that common routes can obtain the path with the most coding opportunities. Simulation results show that the proposed routing protocol can improve network throughput. The application of network coding technology has greatly improved the performance of routing protocol.

The idea of cross-layer can realize the interaction parameters between various protocol layers of wireless Ad hoc network or integrate some network layers, to improve the overall performance of wireless AD hoc network. The cross – layer idea points out a new way for the research of network congestion control algorithm in the future. Xiao Ping's master's thesis [7] proposed an energy-efficient cross-layer energy-saving protocol, which was optimized based on MBCR protocol and considered the retransmission of data packets generated by node use of energy and power. The route consumption is taken as the measurement criterion for route selection, and the cross-layer method is adopted to collect the information such as residual energy and transmitting power of nodes at the physical layer. The transmission power is dynamically adjusted at the link layer according to the requirements of the network layer, and the route is selected according to the total energy consumption from the source node to the destination node at the network layer, thus reducing the consumption of routing protocols. The AODV routing algorithm is optimized. The improved algorithm (CLC-AODV) no longer uses periodic message sending for routing maintenance, but carries out periodic interaction through the grouping of RTS and RSP in the improved cross-layer MAC layer algorithm, obtains the topological relationship between the two hop ranges of nodes, and establishes local response routing. Moreover, efficient routing and fast repair mechanism are implemented to improve the performance of AODV.

Adaptive technology. For multi-hop mobile networks, topology changes degrade network performance. Adaptive technology can automatically select appropriate routing protocols and adjust routing parameters according to network changes to achieve the best results. Adaptive technology is described in this paper. For OLSR active routing protocol, an adaptive adjustment mechanism of protocol parameters and an adaptive multipath routing algorithm are proposed, and an adaptive multipath routing algorithm is proposed based on Linux platform. Moreover, the adaptive adjustment mechanism of OLSR protocol parameters, according to the local node link changes, to adjust the message sending interval. Wang Yanbin [8] studied the on-demand routing protocol AODV and proposed an adaptive routing protocol AODv-AOW combined with clustering algorithm. In the study

of MANET and DTN network architecture in [9], an adaptive routing protocol is proposed, which utilizes the characteristics of different network environments to achieve optimal routing performance. Based on adaptive technology, the routing protocol SEHR improves the transmission performance of wireless AD hoc networks.

Conclusion

Aiming at service quality, energy consumption, security and other problems, a variety of technologies are proposed to optimize the routing protocol, which greatly improves the transmission efficiency of routing protocol in the system. Finding out the way to solve the existing routing problems is a key problem that many scholars have been studying. In this paper, six main technologies are summarized to optimize the routing protocol of wireless AD hoc network. These technologies have significantly improved the performance of routing protocols. With the continuous improvement of people's requirements on network technology, more problems will appear. Therefore, it is necessary to continuously improve wireless AD hoc networks, especially to explore and study the changes of wireless AD hoc networks topology.

References

1. Yang P., Tian C., Zhang L. // Chinese Journal of Applied Sciences. 2006.
2. Sun Y. // Chengdu: Electronics University of Science and Technology. 2016.
3. S Katti., H Rahu.l, Wu H., [et al]. SIGCOMM Computer Communication Review. 2006. 243-254.
4. Zheng S., Hu S., Chen J., Li Z. // Multi-metric wireless based on AODV Research on Routing Algorithm. 2016.
5. S Katti., D Katabi., H Balakrishnan., [et al] // SIGCOMM Computer Communication Review. 2008. P. 401–412.
6. Peng Y. // Chengdu: University of Electronic Science and Technology of China. 2013.
7. Xiao P. // Jilin University. 2011.
8. Wang Y. // University of Electronic Science and Technology of China. 2015.
9. Dong M. // Xidian UniversityScience. 2009.

УДК 004.931

ЭКСПЕРИМЕНТАЛЬНЫЙ ПРОТОТИП ОТКРЫТОЙ СИСТЕМЫ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛЮДЕЙ ПРИ МНОГОКАМЕРНОМ ВИДЕОНАБЛЮДЕНИИ

С.А. ИГНАТЬЕВА, Н.А. ТОМАШЕВИЧ, А.А. ГОЛУБЕНОК, Р.П. БОГУШ

Полоцкий государственный университет имени Евфросинии Полоцкой, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Рассмотрен алгоритм реидентификации людей в распределенной системе видеонаблюдения с использованием сверточных нейронных сетей. Разработан экспериментальный прототип системы реидентификации, позволяющий на кадрах с камер видеонаблюдения выполнять повторную идентификацию людей, формировать набор изображений, упорядоченных в соответствии с идентификаторами для каждого обнаруженного человека. Для обнаружения людей использовалась СНС YOLOv5, для реидентификации применяется DenseNet-121. Выполнена оценка точности работы системы реидентификации в метриках *precision* (точность) и *recall* (полнота).

Ключевые слова: реидентификация, сверточные нейронные сети, информационная система.

Введение

Одной из актуальных задач компьютерного зрения является повторная идентификация людей (реидентификация), предполагающая поиск заданного человека по изображению на кадрах с пространственно-разнесенных камер видеонаблюдения. В зависимости от входных данных выделяют закрытые (close-world) и открытые (open-world) системы. Для открытых систем реидентификации в качестве входных данных используются неразмеченные видео, на которых необходимо обнаруживать людей и динамически формировать галерею изображений, среди которых осуществляется поиск по запросу. Поэтому процесс реидентификации в открытой системе будет состоять из двух этапов. На первом этапе выполняется детектирование людей на изображениях с применением сверточных нейронных сетей (СНС). После того, как человек на кадре видеопоследовательности с одной из камер обнаружен, другая СНС необходима для извлечения отличительных признаков человека, на основе которых осуществляется поиск в галерее. При наличии в ней других изображений этого человека, обнаруженному присваивается соответствующий идентификатор (ID). В противном случае предполагается, что человек впервые попал в поле зрения камеры, и его изображению присваивается новый ID. На этом этапе важное значение будет иметь качество обучения СНС и эффективность извлекаемых ею признаков. В зависимости от поставленной задачи, результатом реидентификации людей может быть отображение всех изображений искомого человека с указанием даты и места их получения или формирование набора данных для всех обнаруженных на видео пешеходов, которые упорядочены согласно их идентификаторам.

Алгоритм повторной идентификации людей

Разработан алгоритм для реидентификации людей, который состоит из основных этапов:
– разделение входных видеоданных на кадры;
– обнаружение людей;
– проверка корректности обнаружений;
– формирование вектора признаков для изображения каждого человека;

– установление соответствия между изображениями людей на кадрах видеопоследовательностей с заданного набора видеокамер;

– присвоение идентификатора человеку на изображении;

– формирование набора изображений обнаруженных и идентифицированных людей (сохранение в файл, располагающийся в папке с соответствующим ID).

На каждом кадре, с использованием СНС выполняется детектирование людей. Действительными считаются те изображения людей, для которых пороговая степень уверенности СНС составляет 60 % и соотношение высоты ограничительного прямоугольника к ширине меньше 2,5, что позволяет не учитывать обнаружения, на которых фигура человека в кадре полностью не отображается.

Для каждого обнаружения с использованием другой СНС вычисляется вектор признаков, отображающий отличительные особенности человека на изображении. Для установления соответствия при реидентификации вычисляется расстояние Эвклида:

$$d_{p,q} = \sum_{i=1}^{n} \left(p_i + q_i \right)^2,$$

где q_i – дескриптор запроса, p_i – дескрипторы изображений ранее обнаруженных людей.

Все изображения группируются и сохраняются в папках, имена которых определяются соответствующими идентификаторами. Для каждого нового изображения человека осуществляется поиск в галерее такого изображения, с которым расстояние между признаками минимально, после чего предполагается, что их идентификаторы одинаковы. Если расстояние между признаками больше установленного порога *Thr*, то считается, что этот человек ранее не присутствовал на кадрах с камеры видеонаблюдения и ему присваивается новый уникальный идентификатор.

Программная реализация прототипа системы реидентификации

Программная реализация разработана на основе представленного алгоритма и состоит из двух модулей: модуль, осуществляющий обработку видео, обнаружение людей и их реидентификацию, реализованный на Python; пользовательский графический интерфейс на C++. В качестве входных данных используются видео, полученные из различных источников (IP-камеры, Web-камеры, видео из файла), которые разбиваются на отдельные кадры с заданным интервалом и приводятся к единому размеру. Для распределения вычислительных ресурсов применяется многопоточная обработка данных. Схема организации потоков представлена на рисунке 1: поток UI (пользовательского интерфейса) и вычислительный поток, осуществляющий все ресурсоемкие вычисления, такие как обработка (захват видеокадров, изменение размера кадра), детектирование людей на изображениях, выделение признаков и непосредственно реидентификация. Оба потока связаны с хранилищем, которое включает два поддерживающих многопоточность контейнера, реализованные в классах AtomicList и AtomicCell.

Первый из них, AtomicList совмещает свойства списка и массива и используется для хранения информации о каждом обнаруженном человеке: его изображение, вектор признаков, идентификатор, информация о номере камеры (видео) и времени получения кадра. Позволяет нескольким потокам читать из него и добавлять новые данные одновременно, требуя при этом меньшего числа переходов по указателям для доступа к нужному элементу по сравнению с обычным списком.

Второй класс AtomicCell используется для передачи текущего кадра видео для вывода на экран и позволяет записывать или считывать только один новый кадр с обнаруженными на нем людьми. AtomicCell содержит счетчик, гарантирующий что кадр не будет прочитан, если между началом и окончанием чтения произошла запись. Для этого перед и после записи счетчик увеличивается на 1. Если в момент начала чтения значение счетчика нечетное, т.е. запись началась, но не завершилась, или не совпадает с показателем счетчика до начала и после окончания чтения, то кадр пропускается.

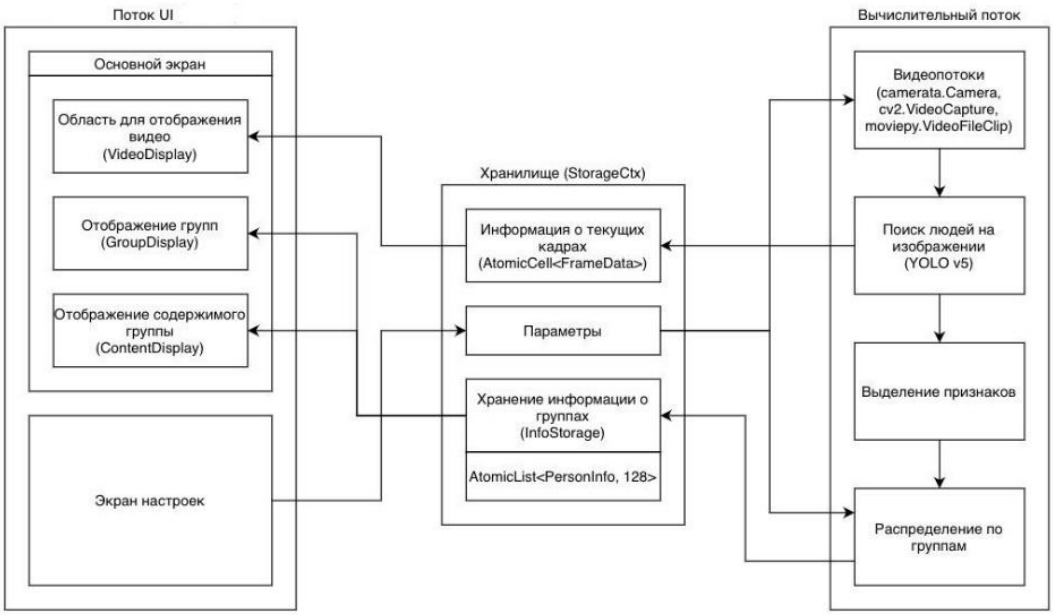


Рис. 1. Схема организации многопоточной обработки

Основной функцией системы является повторная идентификация людей из нескольких видео и формирование набора изображений, разделенных на группы по идентификаторам (*ID*). В качестве источника данных могут выступать видео, полученные с Web- или IP-камер, а также видеоматериалы, полученные ранее с любых других средств видеофиксации и сохраненные в наиболее распространенных форматах: OGV, MP4, MPEG, AVI, MOV, MKV и GIF. Библиотека, используемая для захвата видеопотока, определяется в зависимости от источника видео. При обработке видео из файла для извлечения кадров используется moviepy для Python [1]. Проверка наличия подключенных WEB-камер и захват видеопотока с них выполняется с помощью camerata [2] для Python. Кадры с IP-камер извлекаются с применением библиотеки компьютерного зрения OpenCV [3]. Для передачи видеопотока с IP-камер используется протокол rtsp (real time streaming protocol, потоковый протокол реального времени).

Для обнаружения людей используется СНС версии YOLOv5x6 [4], архитектура которой описана на языке Python и предложена компанией Ultralytics [5]. Для повторной идентификации используется СНС DenseNet-121 [6] в реализации pyTorch и файл весовых коэффициентов net_last.pth [7].

Для увеличения скорости обработки изображений применяется программно-аппаратная архитектура параллельных вычислений для операций с СНС, позволяющая повысить производительность за счет возможностей GPU Nvidia с технологией CUDA. Для ускорения работы на GPU используется библиотека примитивов для нейронных сетей cuDNN, доступная зарегистрированным на сайте Nvidia [8] разработчикам.

Для разработки пользовательского интерфейса используется кроссплатформенная библиотека SFML [9]. Поддерживает распространенные операционные системы (Windows, Linux, MacOS) и большое число языков программирования (C и .Net подобные языки программирования, Java, Python, Ruby, Go и др.). SFML включает 5 основных модулей: System (для управления временем и потоками), Window (для управления окнами и потоками), Graphics (отображение графических примитивов и изображений), Audio (интерфейс для управления звуком) и Network (для сетевых приложений). В процессе разработки прототипа системы реидентификации использовались первые три модуля: System, Window и Graphics.

Интерфейс приложения разбит на два экрана: основной, на котором отображаются видео с обнаруженными людьми, и окно с настройками, вызываемое нажатием клавиши «пробел». Основной экран содержит несколько областей. На рисунке 2 (1) отмечена область из двух чисел в формате *n/m*, где *m* – количество подключенных источников видео, *n* – порядковый номер источника. Стрелки позволяют выбирать какое из видео отображается на экран в области (2) на

рисунке 2. На кадр накладываются прямоугольные рамки, в каждой из которых находится обнаруженный человек. Все люди идентифицируются по внешнему виду и сортируются по группам, согласно их ID. Группы идентифицированных людей располагаются в области (4) на рисунке 2. Изучить содержимое интересующей группы можно нажатием левой кнопки мыши на цель в области 2 в содержимое ограничительного прямоугольника или в области (4) выбрав интересующую группу. На рисунке 2 (5) отображена выбранная группа, а ее содержимое можно изучить в (3). Наведение курсора на интересующее изображение человека выводит в нижнем левом углу экрана (рис.2 (6)) номер источника видео и время получения кадра.

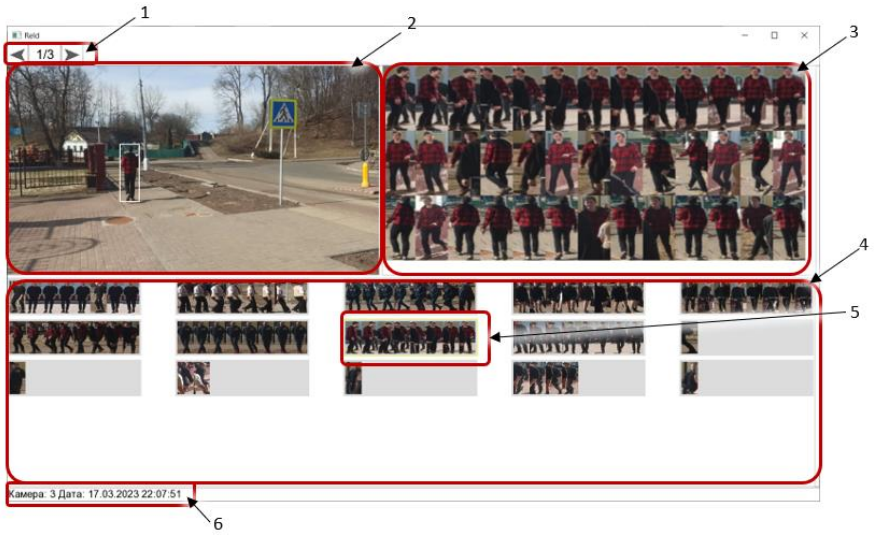


Рис. 2. Интерфейс прототипа системы повторной идентификации человека. Основной экран. 1 – область управления отображением кадров с интересующей камеры; 2 – отображение видео; 3 – содержимое группы изображений искомого человека; 4 – отображение всех идентифицированных людей упорядоченных согласно их ID; 5 – подсветка выбранной группы интересующего человека; 6 – отображение номера камеры и времени получения кадра для изображения, на который наведен курсор мыши в области 3

Все полученные и идентифицированные изображения людей сохраняются в отдельные папки, именами которым служит значение соответствующего идентификатора. В каждой соответствующей папке располагаются все изображения людей, которые предположительно принадлежат одному и тому же человеку. Размер изображений выбирается пользователем из значений [32×64], [64×128] или [128×256] в окне настроек. Кроме выбора размера сохраняемых изображений возможно так же изменить порог, при котором человеку будет присваиваться значение нового ID; добавить источник видео, для чего необходимо указать его расположение (для IP-камеры или видео из файла). Видео с Web-камеры добавляется к обрабатываемым в момент подключения.

Результаты экспериментов

Для определения наиболее эффективного значения порога *Thr*, позволяющего классифицировать изображения людей по идентификаторам проводилось три эксперимента. В первом использовалось три двадцатисекундных видео ролика, на которых присутствовало 5 человек. Во втором – три видео, на каждом из которых 7 человек, длительностью 10 секунд. В третьем эксперименте обрабатывалось три видео для 9 человек, продолжительностью 20 секунд. Примеры кадров для каждого эксперимента представлены на рисунке 3. Результаты приведены в таблице 1. Точность работы алгоритма реидентификации оценивалась по метрикам *precision* (точность) и *recall* (полнота):

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN},$$

где TP – верноположительные предсказания, FP – ложноположительные, FN – ложноотрицательные. *Precision* отражает долю объектов (людей), названных классификатором положительными, и при этом действительно являющихся таковыми, *recall* – отражает какая доля объектов положительного класса из всех верных ответов найдена алгоритмом.

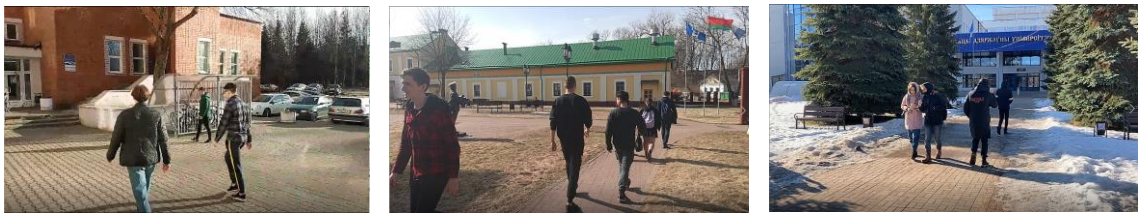


Рис. 3. Кадры из видео, использованные при проведении экспериментов

При расчете метрик не учитывались ошибки детектора. Тестирование выполнялось на персональном компьютере с характеристиками: Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb.

Табл. 1. **Оценка эффективности работы прототипа системы реидентификации**

Особенности эксперимента	Метрика	Значение *Thr*									
		0,98	0,96	0,94	0,92	0,9	0,8	0,7	0,6	0,5	0,4
3 видео по 20 секунд, 5 человек	*Precision*	0,94	0,94	0,94	0,94	0,94	0,94	1	1	1	1
	Recall	**0,92**	**0,92**	0,77	0,77	0,74	0,64	0,47	0,46	0,44	0,39
3 видео по 10 секунд, 7 человек	*Precision*	0,98	0,99	1	1	1	1	1	1	1	1
	Recall	0,94	0,95	**0,96**	**0,96**	0,96	0,93	0,93	0,77	0,67	0,61
3 видео по 20 секунд, 9 человек	*Precision*	0,67	0,78	0,85	0,98	0,98	0,99	0,99	1	1	1
	Recall	0,69	0,77	0,80	**0,90**	0,82	0,79	0,78	0,71	0,65	0,59

Анализ таблицы 1 показывает, что увеличение порогового значения приводит к уменьшению точности в метрике *Precision*. Это связано с тем, что разные люди, имея какие-либо схожие черты, объединяются под одним идентификатором. Чем больше количество человек на видео, тем ниже *Precision* и *Recall* при высоких значениях *Thr*. При этом, если на кадрах небольшое число людей, то увеличение *Thr* позволяет повысить значение *Recall*, т.к. один и тот же человек с разных ракурсов реже оказывается разделен на разные группы по *ID*. Уменьшение *Thr* приводит к понижению показателей точности и полноты. Это связано с тем, что изображениям людей, полученных с разных камер с отличающимися характеристиками, при различных условиях освещенности, ракурсах человека, будут присваиваться новые идентификаторы. Экспериментально установлено, что наиболее эффективен выбор *Thr* в диапазоне от 0,92 до 0,94.

Заключение

Разработанный экспериментальный прототип открытой системы реидентификации позволяет обрабатывать видео с трех камер видеонаблюдения, при присутствии на кадре одновременно до 9 человек. Следует отметить, что количество подключаемых камер ограничивается доступными вычислительными ресурсами. Приложение отображает результат повторной идентификации на экран с указанием даты и места, определяемого по номеру камеры, получения каждого изображения обнаруженного человека, формирует набор данных с идентифицированными людьми. Программное обеспечение может быть использовано также для тестирования алгоритмов реидентификации.

EXPERIMENTAL PROTOTYPE OF OPEN-WORLD PERSON RE-IDENTIFICATION SYSTEM IN MULTICAMERA VIDEO SURVEILLANCE

S.A. IHNATSYEVA, N.A. TOMASHEVICH, A.A. HALUBIONAK, R.P. BOHUSH

Abstract. An algorithm for person re-identification in a distributed video surveillance system using convolutional neural networks is considered. Experimental prototype of re-identification system has been developed that allows people to be detected and re-identification on frames from video surveillance cameras, to form images set sorted according to identifiers for each detected person. CNN YOLOv5 was used for people detection, DenseNet-121 for re-identification. The re-identification system accuracy in *precision* and *recall* metrics was assessed.

Keywords: re-identification, convolutional neural networks, information system.

Список литературы

1. MoviePy. [Электронный ресурс]. URL: https://github.com/Zulko/moviepy.
2. Camerata. [Электронный ресурс]. URL: https://github.com/IntQuant/Camerata.
3. OpenCV. [Электронный ресурс]. URL: https://opencv.org.
4. YOLOv5. [Электронный ресурс]. URL: https://github.com/ultralytics/yolov5.
5. Ultralytics. [Электронный ресурс]. URL: https://ultralytics.com.
6. Huang, G., Liu, Z., Weinberger, K.Q. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). P. 2261-2269.
7. Bohush, R., Ihnatsyeva, S., Ablameyko, S. // Machine Graphics and Vision. 2022. Vol. 31(1/4). P. 93–109.
8. Nvidia. [Электронный ресурс]. URL: https://www.nvidia.com.
9. SFML. [Электронный ресурс]. URL: https://www.sfml-dev.org.

UDC 334.029.3

PHOTOPLETHYSMOGRAPHY AND ACCELEROMETER SENSORS SIGNALS FOR RECOGNIZING PHYSICAL ACTIVITY

S.S. WEI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 17, 2023

Abstract. The utilization of wearable devices to monitor human physiological parameters has been popularized, and due to their low cost, the most common method of monitoring human information in such devices is the use of photoplethysmography (PPG) signals. However, accurate estimation of the PPG signal recorded from the subject's wrist during various physical exercises is often a challenging problem, as the original PPG signal is heavily corrupted by motion artefacts. The article starts with an introduction to how PPG and Accelerometer (ACC) work, and then moves on to the programming, which is then used to provide data processing support for subsequent deep learning by importing data and calculating operations. Long short time memory (LSTM) is built for the paper to recognize activities. The experimental results showed that over 95 % accuracy was achieved in the classification of the test data.

Keywords: Photoplethysmography, Accelerometer, LSTM.

Introduction

The domain of Human Activity Recognition (HAR) has emerged as one of the most popular research topics since the availability, low cost and low energy consumption of sensors and accelerometers, real-time streaming of data, and advances in computer vision, machine learning, artificial intelligence, and the Internet of Things. In HAR, a variety of human activities, such as sitting, standing, walking, running, squatting, and resting, etc., are recognized. Data can be collected from wearable sensors, accelerometers, or images.

PPG is an electro-optical technique, in which the sensor is positioned above the skin and illuminates the skin surface by emitting green light, the sensor receives intensity changes of the reflected light, and the body state is gained through the periodic detection and analysis of the PPG signal. Such non-invasive method of real-time detection on human parameters assumes great practical importance. Numerous studies on the clinical application of photoelectric volumetric pulse waves have shown that the PPG signal contains many human physiological parameters and is an important tool for real-time monitoring of heart rate, blood oxygen saturation, blood pressure, vascular elasticity, etc. The acquisition of PPG signals requires only a special light source and a corresponding sensor, which can be easily integrated into everyday wearable devices to enable continuous monitoring of normal activities without causing discomfort, making PPG signals the preferred choice for health monitoring in everyday life.

However, accurate estimation of the PPG signal recorded on the wrist is often a challenging problem when people wear the wearable device for physical exercise, as the original PPG signal is heavily corrupted by motion artefacts (MAs), mainly due to the relative motion between the PPG source and the wrist skin [1–5]. In order to reduce MAs, a number of signal processing techniques based on data from different sensor types, particularly ACC data, have proven to be very useful [6].

ACC delivers information on the acceleration of the human body during movement. In smartphones and smartwatches, the built-in tri-axial ACC is probably the most common sensor for activity monitoring [7–8]. A combined approach for obtaining PPG and acceleration data is directly available on smartphones and smartwatches devices.

HAR can be regarded as a pattern recognition problem in which machine learning techniques have been proved particularly successful. Various machine learning methods models have been developed for HAR. The primary goal of this paper is to maintain good performance of RNN framework in terms of recognition accuracy, and a RNN was designed for detecting human activities using ACC and PPG four-dimensional data.

Dataset pre-processing

A recent publicly available dataset [9] was used which was from seven different subjects consisting of 105 PPG signals (15 per subject) and a corresponding 105 tri-axial ACC signals sampled at 400 Hz. The seven adult subjects included three males and four females, aged between 20 and 52 years, performing five series of resting, squatting, and stepping activities. The signals were acquired simultaneously and the dataset contained 210 audio clips with a total duration of 17,201 seconds. We use python language for our work.

PPG signals are continually captured during activities from the wrist using Maxim Integrated MAXREFDES100 device. To guarantee a perfect fit of the sensor unit to the skin surface, a specific weightlifting cuff, adjustable by tear-open closure, is used to hold the sensor in place by fully tightening the strap with a cable protruding from the back end of the strap. The PPG signal value is equivalent to the output of an ADC (Analog to Digital Converters) photodetector with a pulse width of 118 μs, a resolution of 16 bits and a full scale of 8192 nA, illuminated by a green LED (Light-Emitting Diode). The ACC signal values on the three axes correspond to MEMS (Micro-ElectroMechanical System) outputs with 10-bit resolution, left-aligned, and a scale of ±2 g.

The primary signal data collected is mixed, ideally regular and stable, and undesirable data is messy and unstable, but has some sort of regular trend, and can be made slightly more regular by filtering the undesirable data to suppress noise. The signal data are sampled according to time and traced to form the time domain data, which is a time referenced regional data.

In a preliminary cleaning step, the following cleaning steps were done on the raw data. If there are occasional spikes or NaN (Not a Number) points in the integrated four-dimensional data, the data from the former five different subjects are used for model training, and the data from the latter two groups of subjects for validation of the model, so only the data from the former five groups are processed.

The PPG signal values for the same subject are highly variable across series and vary considerably over short periods of time within the same series. Normalization allows for better separation of the PPG signal from the motion artefacts, with the following equations:

$$\mathrm{PPG}_{\mathrm{cal}} = \frac{\mathrm{PPG} - \mu_{\mathrm{PPG}}}{\sigma}, \tag{8}$$

$$\mu_{\mathrm{PPG}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{PPG}_i, \tag{9}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{PPG}_i - \mu_{\mathrm{PPG}}\right)^2}, \tag{10}$$

where $\mathrm{PPG}_{\mathrm{cal}}$ is the calculated PPG data, PPG_i means PPG value of the *i*-th data, μ_{PPG} and σ are the average value and standard deviations of the original PPG data, respectively.

The accelerometer is affected by low level noise, the gravitational acceleration is in the three spatial axes of projection, so the data usually has some offset, we remove it by subtracting the average value from the data, the method gives a fine filtering of the signal with the following equation:

$$\mathrm{ACC}_{\mathrm{cal}} = \mathrm{ACC} - \mu_{\mathrm{ACC}}, \tag{11}$$

$$\mu_{\mathrm{ACC}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{ACC}_i, \tag{12}$$

where ACC_{cal} is the calculated ACC data, ACC_i means ACC value of the *i*-th data, μ_{ACC} is the average value of the original ACC data.

The sampling rate of 400 Hz for data can place a significant burden on hardware devices. For the purpose of efficiently downsampling the data, a resampling algorithm that requires a digital filter was not chosen because it would add significant computational cost to the final embedded system implementation. Accordingly, a simple extraction process is implemented in which 1 of the R samples is retained and the rest discarded.

A limited combination of parameters was examined in the vicinity of those already tested, and with a downsampling factor of 10, the best accuracy was achieved when the sample window (before downsampling) was 1200, corresponding 3 seconds with 50 % overlap, and a total of 100 training epochs.

The number of inputs for resting, squatting, and walking activities for the five subjects used for training varies greatly, and the network may end up being biased toward a particular class due to the large difference in numbers. A simple technique to solve this problem is oversampling, a form of data augmentation in which data from less frequent classes are repeated as needed so that the data used for training is more evenly distributed across classes. Oversample only for the first five objects, then the oversampled data are used to train the final network.

Long Short Time Memory Framework

The network model employed in the paper is depicted in Figure 1. It is based on a commonly used architecture for time-based sensor data and is composed of a combination of fully connected layers and LSTM units. The input data consisted of three acceleration axes and PPG, forming a four-dimensional time sequence. The data are then fed into the network in a window of $w \times 4$, with the parameter w being the size of the time point of a single data window.

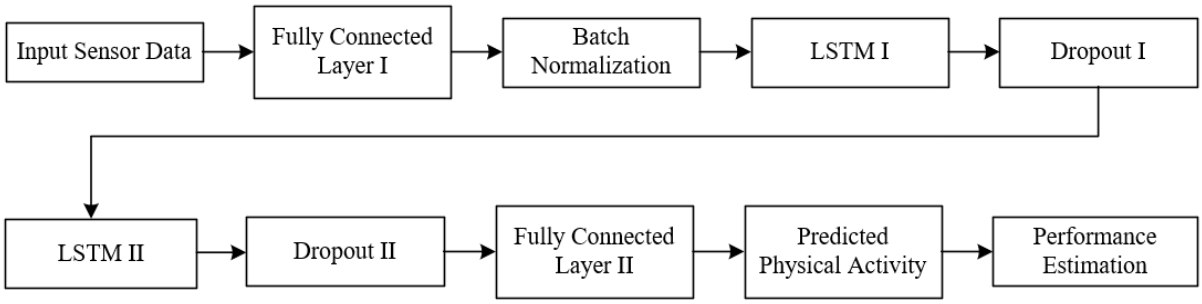


Figure 1. The network model for training dataset

The first layer is fully connected layer and aims to identify the relevant features in the input data. In this layer, the general neuron produces an output value y:

$$y = f\left(\left[w_1, w_2, \ldots, w_n\right]\left[x_1, x_2, \ldots, x_n\right]^T + b\right), \tag{13}$$

where the x_n inputs to the layer and the w_n neuron weights in association with each input, f is the activation function and b is the bias value.

The batch normalization layer, which normalizes the mean and standard deviation of the global data, operates on individual batches of data with training. Then, the recurrent neural network is represented at its core by two cascaded LSTM layers, with the LSTM followed by a dropout layer that randomly discards some of the inputs to reduce overfitting.

In the end, there is a fully connected layer of size 3 which, together with the sparse class cross-entropy loss function assigned to the network, classifies one of these three classes of layers. The loss function represents the error that must be minimized by the training process. The representation of the error varies upon the given function of the network allotted to it. For a categorical cross-entropy function $J(w)$, the error function is as follows:

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log y_i + (1-y_i)\log(1-y_i)], \quad (14)$$

where w is the set of model parameters, N is the number of input test features, y_i and y_i are the true and predicted classes respectively, expressed numerically.

Table 1 shows the details of the individual layers. The RNN, as built in this configuration, has 25,283 trainable parameters.

Table 1. **The parameters of LSTM neural network based on ACC and PGG**

Layer	Input Size	Output Size	Parameters
Fully connected layer I	[w, 4]	[w, 32]	128
Batch Normalization	[w, 32]	[w, 32]	128
LSTM I	[w, 32]	[w, 32]	8320
Dropout I	[w, 32]	[w, 32]	0
LSTM II	[w, 32]	[1, 32]	8320
Dropout II	[1, 32]	[1, 32]	0
Fully connected layer II	[1, 32]	[1, 3]	99

Result and Analysis

It is demonstrated the matrix of confusion that arises when classifying the test data in the same setup in the Figure 2. It is evident that the squat and step activities are the activities with greater error rates, while the rest activity is correctly identified in 99 % of the cases. This may be partly due to the much smaller amount of raw input data for the squat and step activities. Accuracy is the ratio of the sum of true positives (TP) and true negatives (TN) to the total number of records (Num). Figure 3 shows the progress of accuracy (estimated on the training material itself) and loss with respect to the training epochs for the network with no downsampling (original data at 400 Hz). About 100 epochs, the values reach convergence. The accuracy is the evaluation ratio metric to all true assessment results of summarize the total grouping achievement for resting, squatting and stepping activities:

$$\text{Accuracy} = \frac{\text{TP}_{\text{resting}} + \text{TP}_{\text{squatting}} + \text{TP}_{\text{stepping}} + \text{TN}_{\text{resting}} + \text{TN}_{\text{squatting}} + \text{TN}_{\text{stepping}}}{\text{Num}}. \quad (15)$$

In the current setup, a maximum accuracy of 95,36 % was achieved in the test phase for the decimation factor 40. Although dividing the dataset into five training subjects and two test subjects is a natural choice, the limited size of the dataset can lead to biased results depending on the partition chosen.

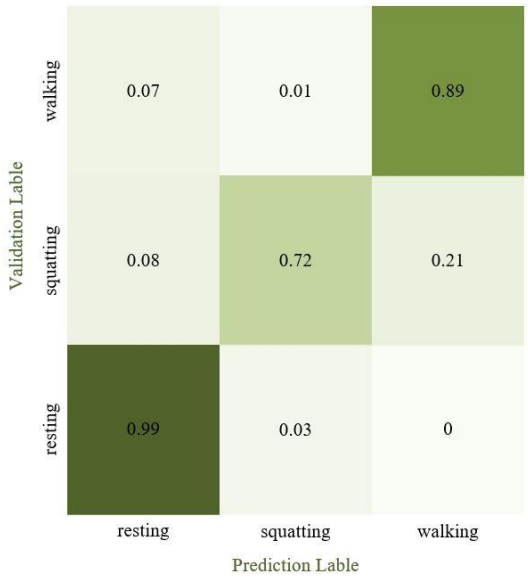


Figure 2. The matrix of confusion for classification

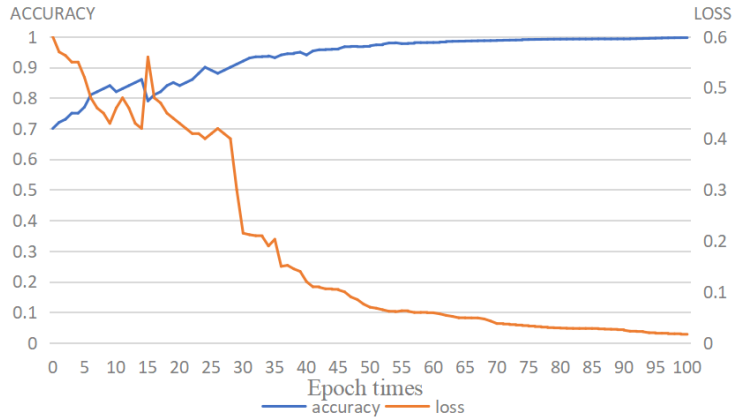


Figure 3. Accuracy and loss progress with respect to training epochs

Conclusion

For recognition, human activity a network model based on LSTM is proposed in the paper. The virgin PPG signal has been severely corrupted by MAs, mainly due to the relative motion between the PPG source and the wrist skin. In order to reduce MAs, the ACC data and PPG were integrated into four-dimensional data, which were processed and analyzed. In an investigation of Python based data analysis of PPG and ACC signals, the LSTM was used for recognizing physical activity. The results revealed that 95,36 % accuracy in classification of the test data was achieved.

References

1. Allen J. // Physiological measurement. 2007. P. 28.
2. Hwang S., Seo J., Jebelli H., [et. al] // Automation in construction. 2016. P. 372–381.
3. Karimian N., Guo Z., Tehranipoor M., [et. al] // Speech and Signal Processing (ICASSP). 2017. P. 4636–4640.
4. Rundo F., Conoci S., Ortis A., [et. al] // Sensors. 2018. P. 405.
5. Abdel-Nasser M., Mahmoud K. // Neural Computing and Applications. 2019. P. 2727–2740.
6. Biagetti G., Crippa P., Falaschetti L., [et. al] // Biomedical Signal Processing and Control. 2019. P. 293–301.
7. Troiano R P., McClain J J., Brychta R J., [et. al] // British journal of sports medicine. 2014. P. 1019–1023.
8. Figo D., Diniz P C., Ferreira D R., [et. al] // Personal and Ubiquitous Computing. 2010. P. 645–662.
9. Biagetti G., Crippa P., Falaschetti L., [et. al] // Data in brief. 2020. P. 105044.

УДК 004.771

ОБМЕН ИНФОРМАЦИЕЙ МЕЖДУ МОБИЛЬНЫМ ПРИЛОЖЕНИЕМ И МИКРОКОНТРОЛЛЕРОМ ЧЕРЕЗ ПРОТОКОЛ MQTT

А.В. ХАРЧЕНКО, В.С. ГАВРИЛЕНКО

Белорусский государственный университет информатики и радиоэлектроники, филиал «Минский радиотехнический колледж», Республика Беларусь

Поступила в редакцию 18 марта 2023

Аннотация. Рассмотрен протокол передачи информации между мобильными устройствами MQTT для передачи данных как между несколькими устройствами на микроконтроллерах, так и для подключения мобильных устройств на микроконтроллерах к смартфону. Показана, схема обмена информацией между клиентом и сервером MQTT протокола. Рассмотрены типы и строение сообщений для работы с протоколом.

Ключевые слова: протокол MQTT, IoT, MQTT сообщение, микроконтроллер, брокер, клиент, сервер.

Введение

Новые интернет технологии стремительно развиваются в современном мире. Широкую популярность, а, следовательно, и востребованность приобретают разработки на базе микроконтроллеров с возможностью управления мобильным приложением через беспроводную связь. На сегодняшний день существует множество различных протоколов, позволяющих подключать мобильные устройства (в том числе на базе микроконтроллеров) к интернету, и соединять их между собой. Одним из таких протоколов является протокол MQTT, который используются в Internet of Things (IoT).

MQTT (Message Queuing Telemetry Transport) – это протокол, сделанный конкретно для IoT. Он предназначен для обмена информацией между разными устройствами и модулями. Отвечает за безопасность соединения, скорость передачи данных и практическое функционирование систем и программ.

Описание принципа работы

Система связи, построенная на MQTT, состоит из сервера-издателя, сервера-брокера и одного или нескольких клиентов. Издатель не требует каких-либо настроек по количеству или расположению подписчиков, получающих сообщения. Кроме того, подписчикам не требуется настройка на конкретного издателя. В системе может быть несколько брокеров, распространяющих сообщения [1]. Схема обмена информации между клиентом и сервером представлена на рис. 1.

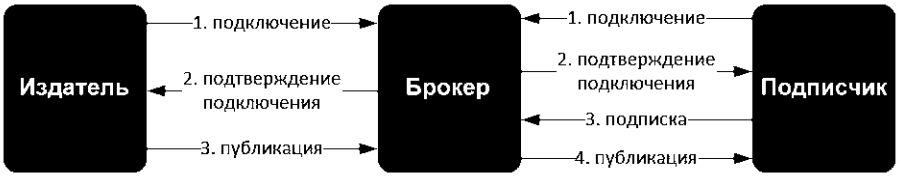


Рис. 1. Схема обмена информации между клиентом и сервером

Есть множество способов настройки клиента для подключения через брокера MQTT. Один из них представлен ниже в виде кода.

```
var options = {
  keepalive: 60,
  username: 'FIRST_HALF_OF_API_KEY',
  password: 'SECOND_HALF_OF_API_KEY',
  port: 8883
};
var client = mqtt.connect('mqtts:mqtt.ably.io', options);
```

Все данные опубликованные или полученные брокером MQTT, будут закодированы в двоичном формате, поскольку MQTT является бинарным протоколом. Это означает, что для получения исходного содержимого нужно интерпретировать сообщение. Ниже представлен способ получение информации с помощью Ably и JavaScript [2].

```
var ably = new Ably.Realtime('REPLACE_WITH_YOUR_API_KEY');
var decoder = new TextDecoder();
var channel = ably.channels.get('input');
channel.subscribe(function(message) {
  var command = decoder.decode(message.data);
});
```

Структура сообщений

Всего в протоколе MQTT существует 15 типов сообщений, которые представлены в табл. 1, где «К» – клиент, а «С» – сервер.

Табл. 1. **Типы сообщений в фиксированном заголовке**

Тип сообщения	Значение	Направление передачи	Описание
Reserved	0000 (0)	нет	Зарезервирован
CONNECT	0001 (1)	К → С	Запрос клиента на подключение к серверу
CONNACK	0010 (2)	К ← С	Подтверждение успешного подключения
PUBLISH	0011 (3)	К ← С, К → С	Публикация сообщения
PUBACK	0100 (04)	К ← С, К → С	Подтверждение публикации
PUBREC	0101 (5)	К ← С, К → С	Публикация получена
PUBREL	0110 (6)	К ← С, К → С	Разрешение на удаление сообщения
PUBCOMP	0111 (7)	К ← С, К → С	Публикация завершена
SUBSCRIBE	1000 (8)	К → С	Запрос на подписку
SUBACK	1001 (9)	К ← С	Запрос на подписку принят
UNSUBSCRIBE	1010 (10)	К → С	Запрос на отписку
UNSUBACK	1011 (11)	К ← С	Запрос на отписку принят
PINGREQ	1100 (12)	К → С	PING запрос
PINGRESP	1101 (13)	К ← С	PING ответ
DISCONNECT	1110 (14)	К → С	Сообщение об отключении от сервера
Reserved	1111 (15)	нет	Зарезервирован

MQTT сообщение состоит из нескольких частей: фиксированный заголовок, переменный заголовок и данные [3].

Фиксированный заголовок присутствует во всех сообщениях. в то время, как переменный заголовок и данные – присутствуют только в определенных сообщениях. Строение фиксированного заголовка представлено на рис. 2.

Рис. 2. Строение фиксированного заголовка

«Message Type» – это тип сообщения. «Flags specific to each MQTT packet» – эти 4 бита отведены под вспомогательные флаги, наличие и состояние которых зависит от типа сообщения. «Remaining Length» – представляет длину текущего сообщения, может занимать от 1 до 4 байта.

Четыре старших бита первого байта фиксированного заголовка отведены под специальные флаги, они изображены на рис. 3.

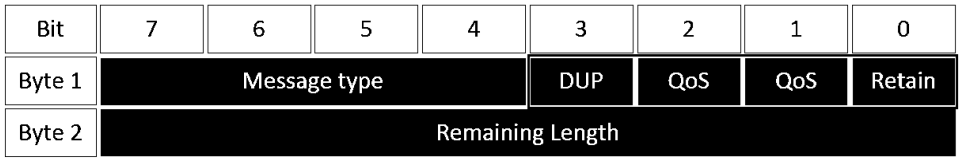


Рис. 3. Расположений страших битов

«DUP» – флаг дубликата устанавливается, когда клиент или MQTT брокер совершает повторную отправку пакета. При установленном флаге переменный заголовок должен содержать Message ID. «QoS»– качество обслуживания. «RETAIN»– при публикации данных с установленным флагом retain, брокер сохранит его. При следующей подписке на этот топик брокер незамедлительно отправит сообщение с этим флагом. Используется только в сообщениях с типом «PUBLISH».

Переменный заголовок содержится не во всех заголовках. В нем помещаются следующие данные. «Packet identifier» – идентификатор пакета, присутствующий во всех типах сообщений, кроме: «CONNECT», «CONNACK», «PUBLISH», «PINGREQ», «PINGRESP», «DISCONNECT». «Protocol name» – название протокола (только в сообщениях типа «CONNECT»). «Protocol version» – версия протокола (только в сообщениях типа «CONNECT»). «Connect flags» – флаги, указывающие на поведение клиента при подключении. Строение переменного заголовка представлено на рис. 4.

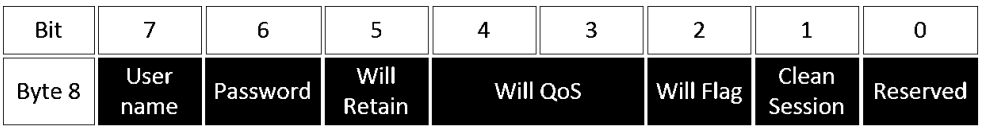


Рис. 4. Строение переменного заголовка

«User name» – при наличии этого флага в «нагрузке» должно быть указано имя пользователя. «Password» – при наличии этого флага в «нагрузке» должен быть указан пароль. «Will Retain» – при установке в 1, брокер хранит у себя «Will Message». «Will QoS» – качество обслуживания для «Will Message», при установленном флаге «Will Flag», «Will QoS» и «Will retain» являются обязательными. «Will Flag» – при установленном флаге, после того, как клиент отключится от брокера без отправки команды «DISCONNECT» (в случаях непредсказуемого обрыва связи, например), брокер оповестит об этом всех подключенных к нему клиентов через «Will Message».

«Clean Session» – необходим для очистки сессии. При установленном «0» брокер сохранит сессию, все подписки клиента, а также передаст ему все сообщения с «QoS1» и «QoS2», которые были получены брокером во время отключения клиента, при его следующем подключении. Соответственно при установленной «1», при повторном подключении клиенту будет необходимо заново подписываться.

Содержание и формат данных, передаваемых в MQTT сообщениях, определяются в приложении. Размер данных может быть вычислен путем вычитания из «Remaining Length» длины переменного заголовка.

Заключение

MQTT предоставляет способ создания иерархии каналов связи – так называемую «ветвь с листьями». Всякий раз, когда у издателя есть новые данные для распространения среди клиентов, сообщение сопровождается примечанием контроля доставки. Клиенты более высокого уровня могут получать каждое сообщение, в то время как клиенты более низкого уровня могут получать

сообщения, относящиеся только к одному или двум базовым каналам, «ответвляющимся» в нижней части иерархии. Это облегчает обмен информацией размером от двух байт до 256 мегабайт. Данный протокол защищает от различных сбоев и неполадок при передачи данных позволяя обмениваться информацией между устройствами, а также выполнять систематизацию локальных сетей в интернете.

EXCHANGE OF INFORMATION BETWEEN THE MOBILE APP AND THE MCU VIA MQTT PROTOCOL

A.V. HARCHENKO, V.S. GAVRILENKO

Abstract. The protocol for transferring information between mobile devices MQTT is considered for data transfer both between several devices on microcontrollers and for connecting mobile devices on microcontrollers to a smartphone. The scheme of information exchange between the client and the server of the MQTT protocol is shown. The types and structure of messages for working with the protocol are considered.

Keywords: MQTT protocol, IoT, MQTT message, microcontroller, broker, client, server.

Список литературы

1. Chen W.J., Gupta R. // Responsive Mobile User Experience Using MQTT and IBM MessageSight, 2014.
2. Boyd B., Gauci J. // Building Real-time Mobile Solutions with MQTT and IBM MessageSight,.2014.
3. Pulver T. // Hands-On Internet of Things with MQTT: Build connected IoT devices with Arduino and MQ Telemetry Transport, 2019

UDC 621.391

DESIGN OF SPEECH RECOGNITION SYSTEM BASED ON ATTENTION MECHANISM

YALU GAO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 20, 2023

Abstract. Attention mechanism is to let the machine pay attention to more key information and ignore the secondary information in complex tasks, that is, to assign different weights to different information to represent different degrees of attention so as to reduce the amount of calculation. The common end-to-end speech recognition model structure is to directly model speech and text, which not only simplifies the speech recognition model but also improves the recognition performance. In order to study the function of attention mechanism in the end-to-end speech recognition system, this paper will take wenet as the basic network framework, and study the conformer part of the encoder and the transformer structure of the decoder in the end-to-end basic framework. These two structures make full use of the attention mechanism.

Keywords: speech recognition, attention mechanism, wenet network, transformer structure, conformer structure.

Introduction

This design will use part of the data in aishell as the data set to train the speech data. By constantly modifying the network framework and changing the attention model, we can get a better recognition rate, so as to get the most appropriate network parameters and achieve a good speech recognition function. In addition, we compare the attention mechanism with other networks, such as CNN, CTC, etc. Through this comparative experiment, we get the shortcomings of attention mechanism itself and its compensation method, such as the combination of attention mechanism and DNN, which makes up for its inability to focus on local features, and the combination of CTC can be used for streaming speech recognition [1].

The main algorithm

Attention mechanism: The attention mechanism is divided into three categories: the first category is the soft attention mechanism, which assigns weight to each input item. The weight is between 0 and 1, because it will be considered, so the amount of calculation is relatively large. The second category is the hard attention mechanism. The weight distribution of each input item is either 0 or 1, so some of them will not pay attention. The advantage is the amount of calculation small, with the disadvantage of possible loss of information. The third type is the self-attention mechanism, and its weight distribution is mainly the input item. We mainly use the self-attention mechanism.

Conformers and Transformers: In the Wenet used in this design, we mainly use two kinds of networks, Transformer and Conformer network [2]. These two networks are typical applications of the attention mechanism, in the part of the encoder network, both can be used. You can choose according to the specific situation, and in the decoder network. For the network part, only the Transformer network can be used. Transformer is composed of multiple Transformer Block groups. As a result, self-attention, res, relu, Feed-Forward layers will be used in each block, as shown in Figure 1. Conformer is composed of multiple Conformer Block composition, each block will use convolutional layer, self-attention layer, residual res, activation relu, Feed-Forward layers [3].

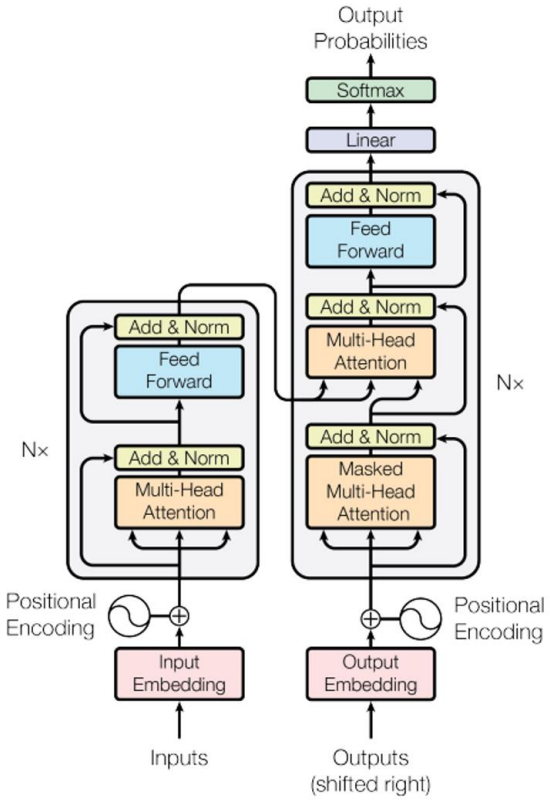


Figure 1. Transformer

Wenet network structure

In order to perform both non-streaming scene recognition and streaming scene speech recognition, Wenet using the U2 model, which is a joint model of CTC and AED. It is composed of Shared Encoder, CTC decoder and Attention decoder [4]. A Shared Encoder can select Transformer or Conformer, the attention decoder is composed of transformer layers, CTC decoder is composed of a Fully Connected Layer and a Softmax Layer. While training at this time, by using the dynamic chunk training method, the Shared Encoder can handle different lengths voice clips. When recognizing, the first step is to go through the CTC decoder to get multiple backups with the highest score. Use the results, and then use the Attention decoder to re-score the candidate results, and choose the highest score final result. When the selected voice segment is infinitely long, it is suitable for non-streaming scenarios that is to get decoding after a complete speech segment, it is also suitable for streaming languages when the selected speech segment is of finite size sound recognition scene [5]. The shared Encoder implements incremental forward operation, and the result of the CTC decoder is also displayed as intermediate results. During decoding, the CTC decoder operates in steaming mode in the first pass, while the attention decoder is used in the second pass to give more accurate results as shown in Figure 2.

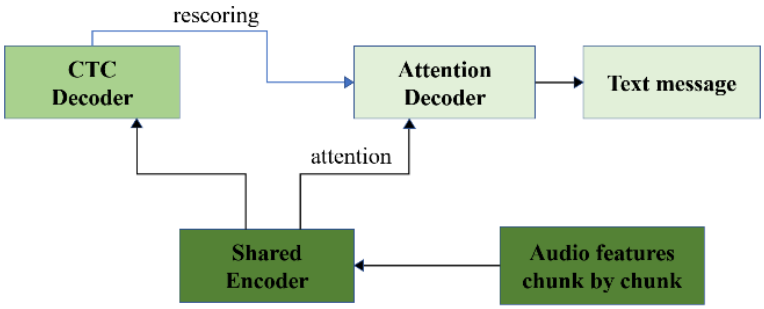


Figure 2. Structure of U2

Comparison of the effects of different decoding methods

This experiment explores two different decoder methods in the non-streaming model, focusing on comparing the real-time rate (RTF) between attention decoder and attention resoring, and the results of the four decoding methods. The full context is used in the model, and the conformer is used for training with a standard convolution kernel size of 15. In AED decoder, $\text{beam} = 10$ is used for decoding; CTC uses prefix beam search to form top n_best hypotheses as reference information for final re-scoring. "/" is used to represent that means it's not important in the process.

Table 1. **Comparison results of different decoding methods**

Decoding method	CTC weight	RTF	CER
Attention decoder	/	0,297	6,02
CTC prefix beam search	/	0,0	6,05
Attention rescoring	0,0	/	5,73
Attention rescoring	0,5	0,182	5,52

Conclusion

The first set of experiments: attention decoder, as shown in Table1. Although the Attention model works well, it still has its own problems. Questions are as follows: 1. Suitable for phrase recognition, poor for long sentence recognition; 2. Training is unstable when noisy data. The second group of experiments: CTC prefix beam search, where the real-time rate of CTC is not the focus of this experiment, so "/" is used to represent it. The third and fourth sets of experiments: attention rescoring, after CTC prefix beam search and attention rescoring, it is found that CTC prefix beam search produces many errors, which can be corrected by the attention rescoring strategy. However, some CTC decoding correct results will also be corrected into errors after attention rescoring, which means that CTC plays an important role in some cases. Therefore, a CTC weight score needs to be added. In addition, tested different CTC weight scores from 0,1 to 0,9, and found that when $\lambda = 0,5,$ it is the most stable.

References

1. Yan Z.J., Huo Q, Xu J. // A scalable approach to using DNN-derived features in GMM-HMM based acoustic modeling for LVCSR[C]. Interspeech 2013. P. 104–108.
2. Yao Z., Wu D., Wang X., [et al]. // Wenet: Production oriented streaming and non-streaming end-to-end speech recognition toolkit[J]. arXiv preprint arXiv. 2021.
3. Vaswani A., Shazeer N., Parmar N., [et al]. // Attention is all you need[J]. Advances in neural information processing systems. 2017. P. 30.
4. Tian Z., Yi J., Tao J., [et al]. // Self-attention transducers for end-to-end speech recognition[J]. arXiv preprint arXiv. 2019.
5. Hannun A. // Sequence modeling with ctc[J]. Distill. 2017. P. 8.

UDC 621.391

SHORT MESSAGE SENDING PLATFORM BASED ON GSM MODEM

ZENG PENG, WEI ZIJIAN, WANG YING

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 20, 2023

Abstract. With the development of smart phones, more and more functions rely on short messages to achieve and Global System for Mobile communication (GSM) is always embedded into a certain system as a tool. The article gives a short introduction about GSM services, and then to realize the SMS sending platform used by Attention command (AT). When correctly connected to the GSM hardware services, it can greatly improve the efficiency of sending messages.

Keywords: GSM, AT command, SMS sending, hardware services.

Introduction

Short Message Service**,** commonly abbreviated as SMS, is a text messaging service component of most telephone, Internet and mobile device systems. With the popularization of mobile phones and smart phones, mobile phone networks have also become increasingly mature. If the mobile phone and the computer network can be combined, it will greatly facilitate people's daily life [1]. However, it is inconvenient to edit short messages on mobile phones, and the input and display are limited, so it is not suitable for industrial applications. Enterprises need an efficient, safe and cost-effective platform. Therefore, the information platform of SMS modem came into being, and the GSM Modem is one of the communication products.

GSM (Global System for Mobile communication) is widely use in the world. Therefore, it is necessary to design a mobile phone short message sending platform based on GSM Modem. With the development of mobile communication technology, GSM network technology is mature and has the characteristics of wide coverage, but the establishment of dedicated data faces many problems such as high cost of transmission network, short communication distance, poor communication effect, etc [2]. Using reasonable and effective GSM network resources, these problems will be solved by using wireless GSM modem.

Over View of development at home and abroad

This is a method that is more suitable for small project development at present. It can be realized only by proficient use of AT commands and serial port programming knowledge, but it needs the support of hardware GSM MODEM. In general, developers will choose to use VC platform and GSM Modem to develop SMS platform [3]. Many technical functions have been completed by using GSM technology at home and abroad:

The GSM remote temperature detection system design of STM32 [4], the remote transmission of data is completed in the form of short messages through the GSM network, and the remote wireless transmission of data is realized through the serial communication RS-232 interface combined with the existing single-chip microcomputer system;

Using the global system of industrial mobile communication (GSM) monitoring of arduino uno [5], after the sensor detects fire and gas leakage, it immediately sends a message through the short message service (SMS) to notify the house owner, firefighter, etc. departments and authorities.

Realize automatic monitoring and short message service system through GSM Modem [6]. It detects unexpected events in the environment, generates alerts with detailed messages, and sends them to users to prevent dangerous situations from happening.

Judging from the functions realized in the above references, GSM is embedded into a certain system as a tool, and when a certain value set in the system is reached, GSM can be used as a sending platform to send warning information to the user's mobile phone conveniently and quickly. In fact, the design completed in this topic is the part of these systems that use GSM to send short messages.

Overview of SMS sending platform

In order to realize the SMS sending platform based on the GSM module, VS2015 is used as the development environment, and C++ is used as the programming language to establish an object-oriented software development platform. The computer uses a serial port or USB interface to connect to the SMS modem through a dedicated connection line, and realizes data communication with the SMS modem through a series of instructions. Through the research of AT command, short message encoding and decoding process, a software program based on dialog box is designed to realize the function of sending short message. Realize the hardware connection between SMS modem and computer by installing the driver. After the computer is successfully connected to the GSM Modem, you can send text messages to the mobile phone conveniently and quickly.

First, use the API communication function to open the serial port, and use the AT command to check whether the CSM module is connected successfully. If the connection fails, the running of the program will end directly, and it needs to be restarted to repeat the above steps until the GSM module is connected normally. If the connection is successful, the content of the short message input by the user is encoded using the above coding method and the PDU format data is generated according to the number of the short message center, the number of the destination user for receiving the short message, the coding method and the validity period of the short message. Analyze the PDU data format of the short message by analyzing the message to be sent stored in the mobile phone [2]. The actual flow chart of the entire sending short message system platform is shown in Figure 1.

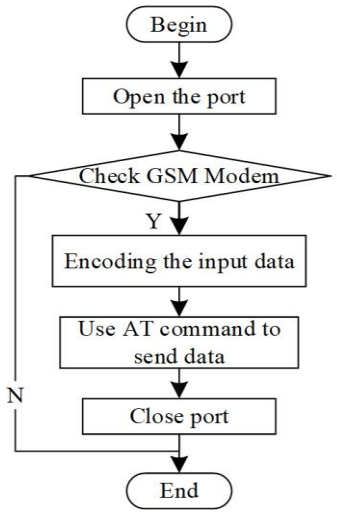


Figure 1. The overall flow of the SMS sending platform

SMS modem hardware interface

First of all, the SMS modem development interface refers to a series of functions or controls provided by the manufacturer of SMS Modem for programmers when programmers program and communicate with SMS Modem. The manufacturer of SMS modem provides 4 kinds of development interface modes for programmers. These four development interface modes are using AT command, SMS modem secondary development kit, SMS modem communication middleware and SMS gateway provided by a third party [3].

Secondly, when users need to obtain relevant information in SMS modem, they need to use AT commands to achieve it. The API function of the SMS modem secondary development kit encapsulates AT commands. The specific hardware connection of SMS Modem is shown in Figure 2. Beginning of the cycle of searching for local single-pixel extrema.

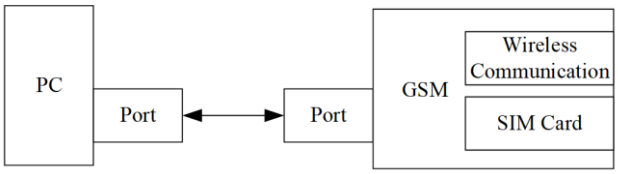


Figure 2. Hardware connection of SMS modem

The hardware interface of SMS modem mainly has two forms: USB and COM serial port. If the serial port mode is used, the hardware connection process is as follows: first, use the serial port connector SMS modem and PC; secondly, insert the SIM card into the communication card slot of the SMS modem, and then connect the external power supply of the SMS modem. If you use the USB connection mode, you only need to insert the SMS modem into the USB port of the computer to realize data communication.

The GSM is a USB serial port single-port industrial type, which supports development kits of multiple programming languages including C++. Its data terminal is always online, and can adapt to low and high temperature working environments. The 2G/3G/4G card of this card does not support other network operators for the time being. Figure 3 shows the SMS modem used in this design.

Figure 3. A kind of industrial-grade USB interface GMS modem

AT command

AT is the abbreviation of the "Attention". The AT command set is an industry standard for the modem communication interface, and it is a command that the modem can recognize and execute. The AT command set is sent from the data terminal equipment (computer) to the terminal adapter (SMS modem). Nokia, Ericsson, Motorola and Hewlett-Packard jointly developed a set of AT command sets for GSM, forming GSMO7.05 and GSMO7.07 technical specifications. These include controls over SMS. The PDU mode is based on the AT instruction set.

Usually, the command of the GSM module starts with the prefix AT, which means "attention". This prefix must be set at the beginning of each command, and at the same time, the end of each command must be terminated with "<CR>" or "<nr>", that is, carriage return. Note that these terminating characters can also be provided by their ASCII values in equivalent hexadecimal, octal, or decimal form. A list of these important commands is given in Table 1 [6].

Table 1. **Command list of AT commands**

Command	Purposes
AT	Check response
AT + COPS?	Read cell operator
ATD "telephone number;"	Dial number
ATH	Hang up command
AT + CMGF = < mode > < CR >	Set messaging mode
AT + CMGS = " telephone number "	Set the target number of calls
< messages > < CR > n032	Send messages

Conclusion

This article introduces the method of using GSM Modem mobile phone text messages, learns AT commands, and finally realizes the sending and group sending of text messages. There are also many functions that need to be perfected. The following features are what are lacking in this design:

First, increase the length of text messages sent by mobile phones. This design supports the sending of text messages with a maximum of 128 characters. If more than 128 characters are used, they need to be sent in batches.

Second, learn the receiving rules of SMS modem and add the function of receiving SMS. In this way, the functions of the software will be further improved, and the receiving and sending modules can be integrated to make it fully functional.

The above problems can be further solved, so that the SMS platform has a certain degree of advancement and universality, and becomes a software product with social value.

References

1. Mingdong T., Junbo. Z., Jianxun. L. // Microcomputer Application. 2007. Vol. 2. P. 74–177.
2. Caiping. X., Hao. W., Guoliang. Z. // Computer Application. 2004. Vol. 5. P. 148–150.
3. Wei. L. Visual C++ network programming case study. Tsinghua University Press, 2013.
4. Yun L., Hansong Z. // Information and Computer. 2021. Vol. 9. P. 14–17.
5. Yadav S., Raghuvanshi R., Soni G., [et al] // IOP Conference Series: Materials Science and Engineering. 2021. Vol. 1136. P. 1–8.
6. Thangarajah A., Wongkaew B., Ekpanyapong M. // International Journal of Computer & Electronics Research. 2014. Vol. 3. P. 2.

УДК 004.032.26

АЛГОРИТМЫ СЛУЧАЙНОГО ПОИСКА В ОБУЧЕНИИ НЕЙРОННЫХ СЕТЕЙ

В.В. МАЦКЕВИЧ

Белорусский государственный университет, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. В работе рассматривается проблема обучения нейронных сетей. Предложены алгоритмы обучения на основе метода отжига и генетического алгоритма. Показано, что разработанные алгоритмы обладают большей эффективностью, чем существующие градиентные методы.

Ключевые слова: метод отжига, генетический алгоритм, метод градиентного спуска, обучение, нейронная сеть.

Введение

В настоящее время происходит стремительное развитие цифровых технологий. Объемы получаемых данных из различных источников стремительно растут [1]. Возникает проблема автоматизации эффективного извлечения полезной информации из больших данных.

Для решения данной проблемы был разработан нейросетевой подход. Проектируется архитектура нейронной сети, настраивается под конкретную предметную область и задачу и применяется для решения задачи. Данный подход является наиболее перспективным из-за универсальности применения нейронных сетей. На практике они способны адаптироваться под любую прикладную задачу, что делает их универсальными. Процесс настройки нейронной сети под прикладную задачу называется обучением. Эффективность полученного решения напрямую зависит от результата обучения [2].

Из-за постоянного роста объема обрабатываемых данных (количества и размерности) постоянно растет размер нейронной сети. Это делает обучение наиболее трудоемким этапом нейросетевой обработки.

Несмотря на достигнутые в данном направлении результаты, проблема обучения по-прежнему является актуальной.

Существуют два основных подхода к обучению нейронных сетей. Один базируется на направленном поиске (перемещение по множеству решений по строго определенному правилу), а другой – на случайном поиске. Наиболее популярными являются градиентные методы (направленный поиск), которые на практике обучают нейронные сети за приемлемое время. Представители другого подхода – методы отжига и генетические алгоритмы, обеспечивают хорошее качество, но работают существенно медленнее.

В работе сравнивается эффективность метода отжига, генетического алгоритма и градиентного спуска для обучения нейронной сети на примере решения задачи сжатия цветных изображений.

Анализ проблемы

Обучение нейронной сети можно сформулировать в виде следующей задачи.

Пусть задана некоторая нейронная сеть N и обучающая выборка, описывающая входные данные задачи X. Пусть задан некоторый функционал качества f описывающий качество полученного решения. В качестве аргументов данная функция принимает нейронную сеть N и выборку X. Данный функционал, как и обучающая выборка, определяются решаемой

прикладной задачей. Необходимо настроить значения параметров сети X таким образом, чтобы достичь минимума функционала f.

Таким образом, обучение нейронной сети является оптимизационной задачей и для нее решения можно использовать все оптимизационные методы.

Наибольшей перспективой обладают приближенные итерационные методы оптимизации. Они универсальны, позволяют регулировать точность полученного решения в зависимости от ограничений на вычислительные ресурсы.

Все итерационные методы оптимизации разделяются на два класса: направленные и ненаправленные методы.

Направленные методы характеризуются наличием строгого правила перехода из рассматриваемого решения в новое. К направленным методам относятся все методы на основе вычисления градиента: метод простого градиента, метод моментов (тяжелого шарика), градиент Нестерова, метод адаптивного момента и т.п. На ранних этапах развития нейронных сетей вычислительные мощности компьютеров были крайне малы. Для обучения сетей был необходим алгоритм с быстрой сходимостью и малым объемом вычислений на отдельной итерации. Любой направленный метод за счет наличия правила переходов существенно ограничивает множество рассматриваемых решений и обладает высокой скоростью сходимости, однако, это приводит к тому, что может быть упущено оптимальное решение. Градиентные методы помимо высокой скорости сходимости, на практике способны получать приемлемое по качеству решение, что обусловило их широкое распространение на начальной стадии развития нейронных сетей и главными методами обучения по сей день. Однако, градиентные методы требуют дифференцируемости целевой функции и могут останавливаться в решениях с около нулевым значением градиента, что приводит к необходимости наложения ограничений на архитектуры нейронных сетей и не всегда близкому к оптимальному решению.

Альтернативным подходом являются ненаправленные методы. В ненаправленных методах на каждой итерации новое решение выбирается случайным образом на основе текущего решения. К алгоритмам случайного поиска относятся генетический алгоритм и алгоритмы отжига. В первом случае на каждой итерации рассматривается конечное множество решений и из него случайным образом генерируется новое множество решений. Во втором случае на каждой итерации рассматривается одно решение и из него случайным образом формируется новое решение. Алгоритмы случайного поиска (ненаправленные методы) не ограничивают множество рассматриваемых решений, что теоретически позволяет достичь оптимального решения, однако из-за того, что множество рассматриваемых решений на практике весьма велико, требуется большое количество итераций для получения решения. Долгое время считалось, что алгоритмы случайного поиска не применимы на практике из-за неприемлемо высокой вычислительной сложности, однако благодаря быстрому росту вычислительных мощностей они способны за приемлемое время получить решение.

Для сравнения двух принципиально различных подходов к решению оптимизационных задач, рассмотрим наиболее сильных представителей из каждого класса. Из направленных методов – метод адаптивного момента, из ненаправленных – метод отжига и генетический алгоритм.

Алгоритм обучения на основе метода отжига

Пусть на конечном множестве допустимых решений Ω определена целевая функция F, и для каждого элемента x множества задано множество соседних элементов $N(x)$. Задачу условной оптимизации в данном случае можно задать в виде тройки (Ω, F, N). Рассмотрим возможности ее решения с помощью метода отжига. Алгоритм включает следующие основные шаги.

Шаг 1. *Предварительный этап*. Инициализация начального состояния нейронной сети $Net_0 = Net(x_{10}, x_{20}, \ldots, x_{m0})$ и последовательности температур $T_0, T_1, \ldots, T_k$, связанных соотношением: $T_k = \dfrac{T_0}{\ln(k+2)}$, $k > 0$, где T_0 – заранее заданное значение.

Шаг 2. *Общая k-я итерация*.

2.1. *Генерация случайных величин*. Генерируется m равномерно распределенных на отрезке от нуля до количества параметров в наборе дискретных случайных величин $a_1, a_2, \ldots, a_m$. Генерируется m случайных перестановок длиной, равной количеству в наборе параметров. Первые $a_1, a_2, \ldots, a_m$ элементов каждой перестановки задают индексы изменяемых параметров в каждом наборе параметров соответственно.

2.2. *Генерация нового решения.* Для каждого изменяемого параметра генерируется равномерно распределенная на отрезке длины l с центром в нуле случайная величина b. Величина l зависит от того, какому набору принадлежит изменяемый параметр и равна $l_1, l_2, \ldots, l_m$ соответственно. Значения l для каждого набора задаются как параметры алгоритма.

Пусть x_i – изменяемый параметр, а его новое значение x_i', тогда: $x_i' = x_i + b$.

2.3. *Принцип перехода.* Пусть x текущее решение, y – новое, сгенерированное на шаге 2 решение. Тогда решение x' на следующей итерации определяется следующим образом: $P(x' = y \mid x) = \min\left\{1, \exp\left(\frac{F(x) - F(y)}{T_k}\right)\right\}$.

2.4. *Критерий останова.* Если время на обучение нейронной сети истекло, то алгоритм завершается. В противном случае производится переход на следующую итерацию.

Как было показано ранее, что предложенный алгоритм обучения на основе метода отжига сходится по вероятности к оптимальному решению, причем из любого начального приближения [3]. Однако, как и большинство алгоритмов случайного поиска, он обладает низкой скоростью сходимости. В экспериментах будет проверена эффективность предложенного алгоритма.

Алгоритм обучения на основе генетического алгоритма

Генетический алгоритм является эвристикой, реализующей идею жадного поиска. Данный алгоритм в зависимости от реализации может быть, как бесполезным, так и крайне эффективным при решении прикладных задач. Однако его главный недостаток – крайне медленная скорость сходимости. Более того, данный алгоритм не гарантирует сходимости к оптимальному решению.

Для проведения экспериментов был разработан следующий вариант генетического алгоритма. Предполагается, что целевую функцию необходимо минимизировать.

Шаг 1. *Предварительный этап*. Генерация нескольких случайных решений. Каждое отдельное решение является полноценной нейронной сетью, архитектура которой задается перед запуском алгоритма обучения и не изменяется. Количество решений N – параметр алгоритма. Для каждого решения вычисляется значение целевой функции.

Шаг 2. *Общая k-я итерация*.

2.1. *«Мутация»*. Из текущего множества решений выбирается решение с наибольшим значением целевой функции. Для выбранного решения производится "мутация" – генерация нового решения из текущего по той же схеме, что и для разработанного алгоритма отжига. Единственное отличие – значения параметров отжига могут отличаться от генетического алгоритма.

2.2. *«Селекция»*. Для полученного решения вычисляется значение целевой функции. Если значение целевой функции для нового решения меньше чем для текущего, то производится замена текущего решения на новое, в противном случае новое решение отбрасывается.

2.3. *«Скрещивание»*. Случайным образом из множества текущих решений выбирается два решения – *a, b*. Для всех значений параметров решений a, b производятся следующие расчеты:

$$\begin{cases} d_i = b_i - a_i \\ c_i = a_i + d_i \cdot \alpha \\ \alpha \in [0; \varphi], 0 < \varphi \le 1 \end{cases},$$

где α – равномерно распределенная случайная величина на отрезке, где φ – параметр алгоритма.

2.4. *«Отбор»*. Для полученного решения вычисляется значение целевой функции. Во множестве решений выбирается решение с наибольшим значением целевой функции. Если значение целевой функции у нового решения меньше, то старое решение заменяется на новое решение, в противном случае новое решение отбрасывается.

2.5. *Критерий останова*. Если время, отведенное на обучение, истекло, алгоритм завершает свою работу, в противном случае производится переход на следующую итерацию.

Из недостатков данного алгоритма также стоит отметить, что, как и любая реализация генетического алгоритма требуется «хорошая» в некотором смысле генерация случайных решений на предварительном шаге. В противном случае скрещивание утрачивает смысл, и алгоритм вырождается в жадный поиск – из одного единственного решения генерируется новое случайное решение путем «мутации» и производится выбор лучшего решения.

Метод градиентного спуска

Метод градиентного спуска заключается в вычислении на каждой итерации градиента целевой функции и изменения значений оптимизируемых параметров в направлении, противоположном градиенту. Методы градиентного спуска обладают высокой скоростью сходимости (не ниже линейной). Например, метод секущих имеет скорость сходимости выше линейной, но ниже квадратической, метод Ньютона имеет квадратическую скорость сходимости. Однако, любой градиентный метод имеет проблему остановки в решениях с околонулевыми значениями градиента. Это могут быть окрестности точек перегиба, локальных минимумов, не совпадающих с глобальным и т.п.

Для решения данной проблемы были разработаны модификации метода: метод моментов (тяжелого шарика), градиент Нестерова метод адаптивного момента [4], метод следования за движущимся лидером [5]. Все модификации градиента заключаются в специфическом правиле пересчета оптимизируемых параметров на основе вычисленных градиентов на различных итерациях. Однако, градиент накладывает ограничения на архитектуру нейронных сетей для решения проблемы затухающего и взрывного градиента и не гарантируют сходимость к оптимальному решению.

В экспериментах для обучения нейронных сетей от направленных методов оптимизации будет использован метод адаптивного момента. Он является одним из наиболее эффективных и обладает стабильностью в качестве полученного решения в отличие от метода следования за движущимся лидером.

Эксперименты

Для сравнения эффективности алгоритмов обучения нейронных сетей на основе различных методов будем решать задачу сжатия изображений на выборке CIFAR-10 [6].

Для экспериментов были выбраны 8-кратное, 16-кратное и 32-кратное сжатия. Более низкие степени сжатия эффективнее производятся с помощью классических алгоритмов сжатия, а более высокие не имеют смысла из-за слишком больших потерь.

Для всех степеней сжатия изображения были разбиты на фрагменты по 4×4 пикселя. Разбиение на меньшие фрагменты приводит к снижению качества сжатия, увеличение, в свою очередь, приводит к слишком большой архитектуре нейронной сети и требует слишком большого объема данных и вычислительных ресурсов для обучения. Каждый отдельный фрагмент сжимает отдельная ограниченная машина Больцмана типа Гаусс-Бернулли. Для 8-кратного сжатия количество нейронов в скрытом слое каждой машины было 48, для 16-кратного – 24, для 32-кратного – 24, но для достижения необходимой степени сжатия был добавлен еще один слой из ограниченных машин Больцмана типа Бернулли-Бернулли с 48 нейронами во входном слое и 24 – в скрытом.

Для обучения ограниченных машин Больцмана градиентным методом будет использоваться алгоритм CD-1. Алгоритм PCD на небольшом числе итераций более эффективный [7], однако, он строится на предположении, что на отдельной итерации параметры обучаемой сети изменяются не существенно, что не соответствует решаемой задаче. Алгоритм CD-k требует в k раз большего объема вычислений на отдельной итерации, чем CD-1 и при этом

достигает лучшего качества [8], однако в решаемой задаче значения градиентов очень велики и использование алгоритма CD-k не целесообразно.

Для обучения первые 8000 изображений были использованы в качестве обучающей выборки, последующие 7000 изображений для валидационной, остальные 45000 изображений сформировали тестовую выборку.

Для оценки эффективности сжатия были выбраны наиболее распространенные функционалы качества MSE (среднеквадратическая ошибка), PSNR (соотношение максимального сигнала к помехе), PSNR-HVS (PSNR с поправкой на особенности человеческого зрения), SSIM (структурная схожесть изображений).

Эксперименты проводились на операционной системе Lubuntu 20.04 с процессором intel i7-4770k, 16 GB 1600MHz оперативной памятью и видеокартой nvidia rtx 3070. Все алгоритмы были реализованы в рамках фреймворка с помощью библиотек OpenCL и OpenMP на языке C++. Фреймворк был настроен на использование видеокарты для обучения.

Время обучения нейронных сетей выбиралось индивидуально для обеспечения наилучшего соотношения качество обучения к затраченному времени (табл. 1).

Табл. 1. **Результат обучения ограниченных машин Больцмана**

Алгоритм обучения	Метод адаптивного момента			Генетический алгоритм			Метод отжига		
Степень сжатия, бит/пиксель	3	1,5	0,75	3	1,5	0,75	3	1,5	0,75
MSE	270	435	917	349	466	798	253	420	670
PSNR	23,9	21,9	18,6	22,8	21,6	19,2	24,2	22,0	20,0
PSNR_HVS	24,2	22,1	18,8	23,0	21,8	19,3	24,3	22,2	20,1
SSIM	0,834	0,765	0,600	0,794	0,747	0,629	0,837	0,762	0,674
Время обучения, ч	1,4	0,52	0,78	2	2	3	4	3	3

Из результатов экспериментов видно, что метод отжига превосходит остальные алгоритмы по качеству полученного решения, однако он примерно втрое медленнее градиентного алгоритма. Обучение нейронных сетей при решении прикладных задач осуществляется лишь однажды, следовательно, трехкратное отставание отжига от градиента по времени обучения не является критическим и является приемлемым.

Генетический алгоритм хоть и превосходит по качеству градиент в сложном случае (обучение ограниченной машины Больцмана типа Бернулли-Бернулли), но уступает методу отжига. Он достигает примерно равного качества с отжигом за счет увеличения времени обучения в десять раз – но это неприемлемо много.

Для быстрого обучения лучшего всего подходит градиентный алгоритм, однако качество полученного решения в таком случае не гарантируется. Более того, по мере роста времени обучения генетический алгоритм и метод отжига продолжают повышать качество полученного решения, в то время как градиент не изменяет качество полученного решения.

Заключение

Экспериментально было показано, что метод градиентного спуска уступает по качеству полученного решения генетическому алгоритму и методу отжига при решении сложных оптимизационных задач. Также экспериментально было показано, что предложенный алгоритм на основе метода отжига является наиболее эффективным алгоритмом обучения, в то время как эффективность генетического алгоритма зависит от конкретной его реализации.

По мере роста вычислительных мощностей компьютеров, метод отжига и генетический алгоритм совершают большее число итераций за фиксированное время, отведенное на обучение, и тем самым повышается качество полученного решения. Для градиентных методов при росте мощностей снижается лишь время обучения, но качество остается неизменным.

Из этого можно сделать вывод, что алгоритмы случайного поиска обладают определенным потенциалом в решении задачи обучения нейронных сетей.

RANDOM SEARCH ALGORITHMS IN NEURAL NETWORKS TRAINING

V.V. MATSKEVICH

Abstract. The paper deals with a neural network training problem. Training algorithms based on the annealing method and the genetic algorithm are proposed. It is shown that the developed algorithms are more efficient than the existing gradient methods.

Keywords: annealing method, genetic algorithm, gradient descent method, training, neural network.

Список литературы

1. Choi Y., El-Khamy M., Lee J. IEEE/CVF International Conference on Computer Vision (ICCV), 2019. P. 3146–3154.
2. You Ya., Li J., Reddi S., [et al]. ICLR, 2020. P. 1–37.
3. Krasnoproshin V.V., Matskevich V.V. // Proc. of the 13-th International Conference "Computer Data Analysis and Modeling". 2022. P. 96–99.
4. Kingma D.P., Ba J.L. // Proc. of the 3rd International Conference on Learning Representations. 2015. P. 1–15.
5. Zheng Sh., Kwok J.T. Proc. of the 34th International Conference on Machine Learning, 2017. Vol. 70. P. 4110–4119.
6. Выборка CIFAR-10 [Электронный ресурс]. URL: https://www.cs.toronto.edu/~kriz/cifar.html.
7. Oswin K., Fischer A., Igel Ch. // Pattern Recognition Letters. 2018. Vol. 102. P. 1–7.
8. Li X., Gao X., Wang Ch. // IEEEACCESS. 2021. Vol. 9. P. 21939–21950.

UDC [004.72+654]-027.31

INFORMATION SYSTEM DESIGN BASED ON MICROSERVICE ARCHITECTURE

HE RUNHAI, LI BOYI, ZHOU QUANHUA, ZHONG WU, ZHANG HENGRUI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 15, 2023

Abstract. In recent years, microservice architecture has become the mainstream approach to modern information system design, and the central idea of this architecture is to improve the scalability, reliability and maintainability of the system by splitting it into small, independent service units. This paper introduces the main ideas and practices of information system design based on microservice architecture, including service splitting, service communication, service governance and service deployment. In addition, this paper analyzes the advantages and disadvantages of the current microservice architecture, and illustrates how to design and implement information systems based on microservice architecture in practical applications with real cases.

Keywords: microservices, architecture design, service communication, service governance, service deployment.

Introduction

As the internet technology advances, information systems have become a vital part of various industries. However, traditional monolithic applications are no longer sufficient for modern applications. This has led to the emergence of microservice architecture as a new approach to software architecture, which is gaining increasing attention. Microservice architecture splits a system into small, independent service units, which enhances system scalability, reliability, and maintainability [1]. This paper aims to introduce the main ideas and practices of designing information systems based on microservice architecture. We will also analyze the advantages and disadvantages of microservice architecture and provide real-life examples of how to design and implement information systems based on microservice architecture in practical applications.

Service Splitting

In microservice architecture, splitting the entire system into multiple small, independent service units is a very critical step. The purpose of service unbundling is to break down a large, complex system into multiple, small, easily managed and maintained parts. The principle of unbundling is usually to divide parts with clear responsibilities and functions into separate service units [2]. For example, an online shopping system can be split into multiple services, such as catalog service, cart service, order service, payment service, etc.; the division is show in the Figure 1.

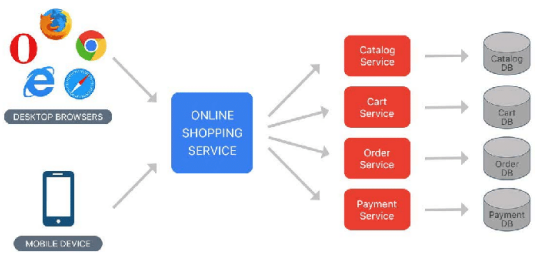


Figure 1. Service Splitting

Service Communication

In a microservice architecture, communication between services is achieved through the network. There are usually two ways of communication between services: synchronous and asynchronous. In synchronous communication, services need to wait for each other to respond, while in asynchronous communication, services do not need to wait for each other to respond. The differences between these two ways are shown in the Figure 2. The advantage of synchronous communication is that it is simple and straightforward, easy to understand and debug, but it can cause performance bottlenecks in large-scale systems. Therefore, asynchronous communication is usually used in large-scale systems, such as message queues, event-driven, etc. With asynchronous communication methods, services can hand over requests to other services for processing and respond when they get the results after the processing is completed. Service communication also needs to consider the protocol, format, interface and security aspects between services, so it is necessary to know about network communication, API design, security and data format.

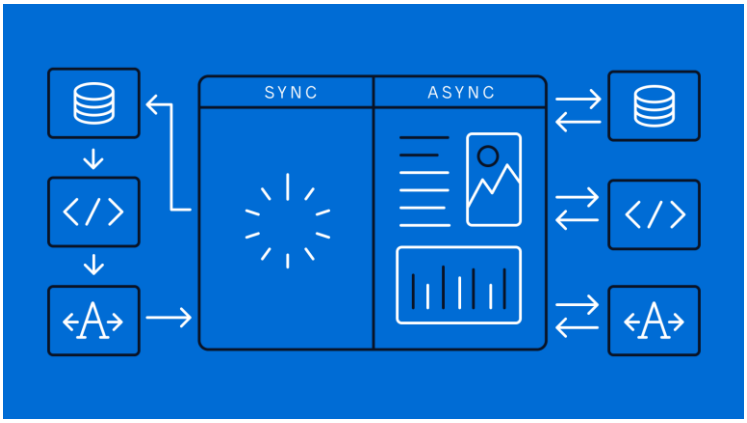


Figure 2. The difference between synchronous and asynchronous

Service Governance

Services in microservice architectures are often decentralized, so the governance and management of services is a very important issue. Service governance includes aspects such as service registration, service discovery, load balancing, failover and monitoring [3].

Service registration and discovery refers to registering services to service centers so that other services can discover them and communicate, there are some common SpringBoot combined with microservice architecture service discovery tools such as Eureka, Consul, Zookeeper, etc. These service discovery tools and frameworks provide APIs and UI to manage service registration and discovery, making it easy for developers to use them to build microservice architecture.

Load balancing is the balanced distribution of requests to multiple service instances to improve system performance and availability, there are some common SpringBoot combined with microservice architecture load balancing are Ribbon, Eureka, Feign, Zuul, etc. Failover is the ability to automatically switch to another available instance when a service instance fails. Monitoring is the ability to improve service availability and performance by monitoring the operational status and performance metrics of the service. The implementation of service governance requires learning about service registration and discovery, load balancing, failover, and monitoring.

Service Deployment

Services in a microservice architecture are usually deployed independently, so the deployment and management of services is also an important issue. The deployment of services needs to consider several aspects, such as runtime environment, configuration management, containerization, etc. Among them, containerization is a common deployment method that packages services and their dependencies into container images for deployment in different environments. Containerization can also improve the

portability, scalability and reliability of services. Service deployment requires knowing about containerization techniques, automated deployment and configuration management.

Advantages and disadvantages analysis

A complete microservice application development and deployment is divided into the process of technical framework building, basic function development, business function development, multi-system integration, production environment deployment and upgrade, production environment problem solving, problem handling, etc. [4]. The development scenarios of directly using the original technology for microservice application development and using the microservice application development platform are different.

Microservices architecture has many advantages, for example, it can improve the scalability, reliability and maintainability of the system, the development scenario of using microservice application development platform shields the underlying complex technology, simplifies the development process, avoids a lot of repetitive development work, and reduces the development difficulty, improves efficiency and quality, increases the productivity of the team. However, microservice architecture also has some disadvantages, for example, it increases the complexity of the system, requires higher technical level and management cost, and may have communication problems between services. Therefore, you need to weigh the advantages and disadvantages when choosing a microservice architecture, and make a choice based on specific business needs and technical strengths.

Practical Case Analysis

Uber is a leading global mobility service platform that provides a wide range of mobility services such as online cars, cabs, carpooling, bike sharing and flying cars. To support the efficient and rapid growth of these services, Uber has adopted a microservices architecture to build its information system. Uber’s microservices architecture consists of thousands of microservices that can be deployed, extended and upgraded independently.

Each microservice has its own specific responsibilities and functions, such as passenger information management, driver routing, payment services, and more. This architecture allows Uber's information systems to be more flexible, scalable and maintainable. In Uber's microservice architecture, each microservice has its own database and API, and the microservices communicate with each other through RESTful APIs, which allows for loose coupling and distributed deployment. In addition, Uber employs distributed message queues to support asynchronous communication between microservices [5]. These technologies allow Uber's microservices to scale horizontally to meet massive concurrency and high load requirements. To support the microservice architecture, Uber also employs containerization technologies and automated deployment tools. Uber uses Docker as the containerization technology to make the deployment and management of microservices easier and more efficient. Also, Uber has developed its own automated deployment tools, such as DeployBot and Cherami, to automate the deployment and management of microservices. Uber also uses a variety of monitoring and troubleshooting tools to ensure high availability and performance of microservices [5]. For example, Uber uses Zipkin to track calls and response times between microservices, Hystrix for circuit breaker patterns and failover, and Prometheus to collect and analyze system metrics. These tools help Uber quickly diagnose and resolve microservice failures and performance issues.

Overall, Uber's microservices architecture is a highly scalable, loosely coupled and distributed architecture that can support rapid iteration and innovation across multiple travel services. At the same time, Uber uses a variety of technologies and tools to ensure high availability and performance of microservices. This architecture has proven to be very effective in practice and can provide useful references and lessons for other enterprises.

Conclusion

Microservice architecture has emerged as a new approach to software architecture that enhances the scalability, reliability, and maintainability of information systems. The splitting of services, communication, governance, and deployment are all critical aspects of designing information systems

based on microservice architecture. Microservices architecture offers many advantages, such as scalability, reliability, and maintainability, but it also has some disadvantages, such as increased complexity and higher management costs. Therefore, the choice to adopt microservice architecture should be based on specific business needs and technical strengths. As exemplified by Uber, microservices architecture has been implemented successfully in practical applications, enabling efficient and rapid growth of services, and improving flexibility, scalability, and maintainability of information systems.

References

1. Fowler, M. Microservices: a definition of this new architectural term, 2014. [Electronic resource]. URL: http://martinfowler.com/articles/microservices.html.
2. Xiao B, Wang YF. / An efficient microservice application development platform architecture design. // China Science and Technology Information. 2023. P. 96–98.
3. Zhao, C., Zhang, J., Zhang, L., C. Research on key technologies of microservice architecture. // Journal of Software. 2017. Vol. 28. P. 1235–1253.
4. Zhang, J., Zhang, C. / A Survey of Microservice Architecture: Principles, Characteristics, and Technologies. // ACM Computing Surveys (CSUR). 2018. Vol. 51. P. 1–35.
5. Uber Engineering. How Uber's microservices platform is built, 2016. [Electronic resource]. URL: https://eng.uber.com/microservice-architecture/.

UDC 004.62

RESEARCH ON TEXTURE IMAGE FEATURE EXTRACTION METHOD

J.K. CHEN, J.X. FU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 22, 2023

Abstract. In this paper, we give several classical feature extraction methods, including grayscale co-generation matrix, Gabor and wavelet transform features, and local binary pattern series features. We introduce the basic principles of these feature extraction algorithms and some derivative methods respectively. Finally, we analyze the advantages and disadvantages of the existing feature extraction methods: grayscale covariance matrix can analyze the arrangement rules of image texture and extract local spatial features of the image, filtering methods and local feature extraction methods are widely used, but the extracted features do not provide a good description of the image structure; and the multi-feature fusion operation brings huge computational effort. Therefore, the future developable directions are proposed based on the existing problems and difficulties in processing texture images.

Keywords: texture image segmentation, feature extraction, image processing.

Introduction

Image, as a visual description of things, has many attributes. Such as chromaticity, brightness, saturation, etc., where texture is an important property of an image. Texture mainly represents the structural features of physical surfaces, which are complex and have many properties. Texture has a basic unit called texture primitive, and texture is a structure composed of a large number of texture primitives arranged according to a given law. Because each texture is arranged differently, the texture of each image is different and has variability. Textures are mainly classified into natural and artificial textures, and almost all images have different texture information.

And the image texture feature extraction is a very important part in image texture classification [1], texture segmentation, texture synthesis, etc. A good texture feature should have the advantages of small computational effort, small feature dimension and strong differentiation ability. In this paper, we introduce several existing feature extraction methods for texture images and analyze the problems and improvement aspects of each method.

Gray Level Co-occurrence Matrix

The texture of an image is formed by the recurrence of gray level distribution in spatial locations, so there is spatial correlation between different pixel gray levels. The Gray Level Co-occurrence Matrix (GLCM) is a classical texture feature matrix, which is computed in the spatial domain based on the following assumptions: The texture information in image I is contained in the overall or average spatial relationship between the gray levels in the image and each other.

The grayscale co-generation matrix is a calculation of four closely related joint and conditional probability density functions, and these four calculated values represent the texture features of the image. Among them, the second-order joint conditional probability density function $P(i, j, d, \theta)$ represents the number of occurrences of the combination of gray levels i and j in an image for two pixels with gray levels i and j, under the condition that the distance is d and the directional angles differ by θ. For an image with one gray level, GLCM is a G×G matrix. For an image of size 4×4 with four gray levels (0 to 3), as shown in Figure 1, *a*. The general form of the image grayscale covariance matrix is given in

figure 1. According to Figure 1, *b*, it can be seen that for an image with four gray levels, its grayscale covariance matrix is a 4×4 matrix, and the grayscale covariance matrix varies at different angles. In the calculation of the grayscale covariance matrix, the distance d and the angle θ are two important parameters, and in general, the value of d is taken as 1. The four plots in figure 2 represent the values of probability density functions under the conditions of distance d as 1 and θ as 0°, 45°, 90°, 135°, respectively.

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

a

	0	1	2	3
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

b

Figure 1. 4×4 image with four gray-level values and general form of any GLCM for image with value 0-3: *a* – 4 gray level images; *b* – General form of GLCM

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

6	0	2	0
0	4	2	0
2	2	2	2
0	0	2	0

2	1	3	0
1	2	1	0
3	1	0	2
0	0	2	0

4	1	0	0
1	2	2	0
0	2	4	1
0	0	1	0

Figure 2. Four coeval matrices with distance 1

When the gray level G is relatively large, the size of the gray co-occurrence matrix will be very large, which makes the subsequent computation increase dramatically. Therefore, we can analyze the histogram of the texture image, perform appropriate grayscale transformation to achieve the purpose of compressing the gray level without affecting the texture quality as much as possible, and then calculate the grayscale co-occurrence matrix.

Image filtering method

Tuceryan and Jain summarized five major categories of texture features [2], which are: statistical-based, geometric-based, structure-based, model-based, and signal processing-based features. Image filtering methods mainly extract signal processing-based features, mainly including Laws texture template, ring, wedge filter, binary Gabor filter, wavelet transform, discrete cosine transform, optimized Gabor filter, optimized finite impulse response filter, etc. These methods extract the local energy of filter response as features. It has been proved by previous experiments that these methods are more effective than statistical and model-based methods for segmentation. The basic assumption of most filtering methods is that the distribution of energy in the frequency domain identifies textures, so that the spectral energy features of different textures are different if the frequency domain of a texture image is decomposed into a sufficient number of subbands. We present two common methods in the following:

1. Gabor Filters.

Jain and Farrokhnia proposed a set of Gabor filters (also known as Gaussian-shaped bandpass filters) for binary coverage of the radial spatial frequency range and multiple directions, and the designed

binary filter bank captures image texture information well due to the joint optimal resolution of the filters in time and frequency. The basic even-symmetric Gabor filter is a band-pass filter with unit impulse response in the 0° direction:

$$h(k,l)=e^{\frac{1}{2}\left(\frac{k^2}{\sigma_x^2}+\frac{l^2}{\sigma_y^2}\right)}\cos\left(2\pi f_0 k\right) \tag{1}$$

In equation (1), f_0 is the radial center frequency, (k, l) is the reference coordinate system, the filter response in other directions can be obtained by rotating the reference coordinate system (k, l). The filter has an infinite unit impulse response, but in practical experiments it is approximated as a finite length filter. Jain et al. used five radial frequencies and four directions for an image of size 256 × 256, where the discrete radial center frequencies are $\frac{\sqrt{6}}{2^6},\frac{\sqrt{6}}{2^5},\frac{\sqrt{6}}{2^4},\frac{\sqrt{6}}{2^3},\frac{\sqrt{6}}{2^2}$, angles of 0°, 45°, 90°, 135°.

2. Wavelet transforms.

Wavelet transforms include discrete wavelet transforms, orthogonal wavelet transforms, two-dimensional wavelet transforms, and wavelet packet transforms. Transforms like discrete wavelet correspond to critical sampling filter banks with specific filter parameters and sub-band decomposition; therefore, wavelet transform methods are filter bank methods. The application of wavelet transforms and its derivatives to texture image recognition has received extensive attention in the related literature, where Mallat applies the standard discrete wavelet transform to feature extraction, i.e., critical extraction with a binary subband structure. All wavelet transforms are obtained by stretching and translating the wavelet basis function. The difference between wavelet and Fourier transform is that: the wavelet transform replaces the infinitely long trigonometric basis with a finite length wavelet basis that decays. The wavelet formula is:

$$WT(a,\tau)=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty}f(t)\varphi\left[\frac{t-\tau}{a}\right]dt \tag{2}$$

There are two variables in equation (2): scale an and translation t. Scale a controls the scaling of the wavelet function and translation t controls the translation of the wavelet function. Each wavelet transform has a mother wavelet and a father wavelet, and the basis function of the wavelet transform is formed by the scaling and translation of the mother wavelet and the father wavelet. In general, the scaling multiplier is of the order of 2, and the size of the translation is related to the degree of scaling. The basic functions derived from different mother and father wavelets are also different. In digital image processing, the discrete wavelet transform and discrete wavelet packet transform are strictly sampled multirate filter banks that decompose the signal at different scales, for which the choice of scale is determined according to different situations, to obtain low-frequency information and high-frequency information. Among them, the low-frequency information is important and contains the signal characteristics, while the high-frequency information contains the details and differences of the signal. However, strictly sampled filter banks usually imply inaccurate texture edge localization. Later, related researchers have used a complete wavelet representation, i.e., wavelet frames, to alleviate this problem and improve the final results.

Local Binary Patterns

Local Binary Patterns (LBP) was first proposed as an efficient texture description operator, and it has been widely used due to its excellent ability to depict local texture features in images. They are invariant to monotonic grayscale variations. The image is scanned line by line, and for each pixel point in the image, the grayscale of the point is used as a threshold to binarize its surrounding 3×3 8-neighborhoods, and the result of the binarization is formed into an 8-bit binary number in a certain order, and the value of this binary number (0 to 255) is used as the response of the point. Based on the traditional LBP, many variants of the method have been proposed, among which the more famous ones are the grayscale and rotation invariant binary patterns proposed by Ojala et al. in 2002 [3, 4]. Subsequently, the local ternary mode LTP was also used to describe the texture structure by encoding the differences between pixels into three classes. Liao et al. then suggested to use the most frequently

occurring mode [5], called the principal local binary mode DLBP, as texture features, however, only the principal mode frequencies were considered in DLBP, while the types of information modes were discarded. The features obtained by the LBP family of methods are a local descriptor, which cannot capture the larger scale texture structure features. Therefore, patch sampling and geometric sampling structures have been proposed again to encode the structural information. To address the noise sensitivity of LBP, many approaches have been tried, including the use of local averaging, frequency or transform domain components, or error correction mechanisms to mitigate this problem. In addition, alternative coding rules have been designed for specific applications, such as the complete local binary counting method CLBC for texture classification and the local directional digit LDN for facial recognition. Combining information from different aspects of an image to generate a histogram can lead to a feature representation with strong discriminative power, for example, Guo et al. proposed the complete local binary pattern CLBP by encoding three complementary components, i.e., the central pixel as well as the sign and size of the local differences [6, 7]. This method adds the central pixel and the size value of the local differences compared with the traditional LBP method, and the information from the three aspects form a more complete feature histogram that contains more local information of the texture. Compared with LBP, the ability to recognize images is much better. In addition, there is also the use of locally enhanced binary code LEBC to enhance the binary code LBP for texture representation and good results were obtained. Other works include applying the idea of LBP to encode the neighborhood information of Gabor features and single gene signal features to enhance face recognition. These algorithms also enhance the discriminative power of the algorithm by expanding the LBP codes.

Conclusion and Outlook

From the introduction of the methods above, it is clear that there are many methods for feature extraction. Grayscale co-occurrence matrix can analyze the arrangement rules of image texture and extract local spatial features of the image. In addition, there are filtering methods as well as local feature extraction methods. In fact, most of the texture segmentation methods extract features from local image patches and then provide them to general clustering or segmentation model algorithms. Various descriptors have been devised by various researchers to characterize texture appearance, and widely used filtering operations are based on filters and statistical models, where filters are used to decompose an image into a set of sub-bands using filter banks, and the model attributes texture to some underlying probability distribution. Although these features can describe the image well, a single texture descriptor alone does not describe the structure of the image well enough and tends to ignore the important information of the image. The simple multi-feature fusion operation, in turn, leads to high feature dimensionality, which causes high computational complexity and expensive costs.

Recent work in texture analysis has shown that texture descriptors constructed by convolving an image with a bunch of filters, based on the local distribution of filter responses, show promising texture recognition performance. Such descriptors can then be combined with well-established segmentation methods to segment texture images. However, there are two main problems with this processing: The first problem stems from the high feature dimensionality of multiple filter responses and their distributional representations. Many widely used segmentation methods, such as graph segmentation methods, curve evolution, and mean shift, rely heavily on measuring the distance between local features, and thus applying them to the distance calculation operation of texture descriptors requires high computational cost. Moreover, choosing the appropriate distance metric for a high-dimensional space is always a tricky problem. Although dimensionality reduction techniques can be used, the suitability of the technique for features usually lacks theoretical justification support. Contending with the above problems, possible improvement aspects are: optimizing and proposing some image segmentation algorithms based on feature extraction for alleviating the current dilemma faced; proposing new segmentation models and algorithms that filter out redundant information in the ensemble of signs and extract a CMI set of features that better describes the image.

References

1. Wang L. [et al.] // Signal Processing, 2018, P. 27–35.
2. Jain A., Farrokhnia F. // Pattern Recognition, 1991, Vol. 24. P. 1167–1186.
3. Ojala T., Pietikainen M. // Pattern Recognition, 1996, Vol. 29. P. 51–59.
4. Ojala T. [et al.] // IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, Vol. 24. P. 971–984.
5. Liao S., Law M., Chung A. // IEEE Transs Image Proces, 2009, Vol. 18. P. 1107–1118.
6. Sengur A., Guo Y. // Computer Vision and Image Understanding, 2011, Vol. 115. P. 1134–1144.
7. Guo Z., Zhang L., Zhang D. // IEEE Transactions on Image Processing, 2010, Vol. 19. P. 1657–1663.

UDC 334.029.3

DEVELOPMENT OF RECURRENT NEURAL NETWORKS

S.S. WEI

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 17, 2023

Abstract. The Recurrent neural network (RNN) has been the main implementation of neural network sequence model, which is the standard processing tool for machine translation, machine question answering, and sequence video analysis, as well as the mainstream modeling tool for problems such as automatic handwriting synthesis, speech processing, and image generation. In the paper, the traditional RNN and the improved Long short-term memory (LSTM) are described in detail.

Keywords: RNN, neural network sequence model, LSTM.

Introduction

Artificial neural networks (ANNs) are algorithms inspired by the neuroscience of the brain, and their basic building blocks are neurons [1]. The nervous system of the brain is the most complex organism in the human body. It consists of nearly 86 billion neurons, each with thousands of synapses connected to other neurons, and together these countless neurons form the brain's complex nervous system. Artificial neurons are abstractly constructed by simulating the functions of biological neurons in the brain, receiving a series of signal inputs and activating them to produce corresponding outputs, and establishing connections between them according to a specific topological network structure.

In the traditional neural network model, it is fully connected from the input layer to the implicit layer to the output layer, and the nodes between each layer are connectionless. However, there is a class of problems that traditional neural networks cannot solve, that is, the training sample input is a continuous sequence, and the length of the sequence varies, such as time-based sequences: a continuous segment of speech, a continuous segment of handwritten text. These sequences are long and of different lengths, so it is difficult to split them into individual samples for training by traditional neural network models. Therefore, a new neural network structure is needed, and thus a recurrent neural network is proposed.

Recurrent neural network is a kind of network model that can accurately model the temporal data. Similar to convolutional neural networks which are good at processing and learning from gridded data, convolutional neural networks are able to process gridded data of variable size, most recurrent neural networks are also able to process sequential data of uncertain length [2-4]. Currently recurrent neural networks have been applied to several fields, especially when there is temporal correlation in the data. The specific structure of recurrent neural networks is that the recurrent neural network will store the memory of previous information in the hidden layer and then input to the currently computed hidden layer unit, the internal nodes of the hidden layer are no longer independent of each other, but have messages to each other.

However, RNN networks still have some shortcomings, such as gradient disappearance and gradient explosion problems. These two situations are usually caused by the backpropagation algorithm. If the gradient disappears, the network layer weights cannot be updated and the training is terminated early, while the gradient explosion causes the network layer parameters to change too much at one-time step and the learning process is unstable. The most intuitive impact of these two problems is that RNNs cannot effectively utilize historical information and are difficult to achieve long-term dependence. The LSTM solves this problem by introducing a gating mechanism to better control the speed of information propagation and thus change the backpropagation structure.

Feedforward neural network

Feedforward neural network (FNN), is a forward-structured neural network. Each neuron is arranged in layers, and each neuron is connected to the neuron of the previous layer only. The output of the previous layer is received and output to the next layer, and there is no feedback between the layers, which is one-way propagation. In this neural network, each neuron can receive signals from the neurons in the previous layer and produce outputs to the next layer. Layer zero is known as the input layer, the last layer is referred to as the output layer, and the other intermediate layers are called hidden layers. The hidden layer can be one layer or multiple layers. Figure 1 shows the structure of FNN. Each node except the input node is a neuron with a nonlinear activation function, which defines a $f(x,\theta)$ mapping that enables the network to achieve the best approximation of the function by learning the value of the parameter θ.

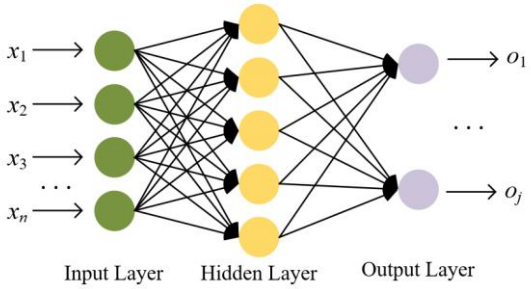


Figure 1. The structure of FNN

A neural network is a nonlinear combination of weighted inputs through an activation function that forms a nonlinear decision boundary, and if the inputs are relevant to the neural network, they are selected for activation, stored, and passed backward. If they are not important, they are suppressed and the inputs do not continue to be passed backwards. The activation functions for neural networks are Tanh, Sigmoid and Relu activation functions, as shown in Table 1.

Table 1. **Three common expressions for activation functions in neural networks**

Activation functions	Function expressions	Derivative expressions
Tanh	$f(x)=\frac{1-e^{-2x}}{1+e^{-2x}}$	$f'(x)=1-f(x)^2$
Sigmoid	$f(x)=\frac{1}{1+e^{-x}}$	$f'(x)=\frac{e^{-x}}{(1+e^{-x})^2}$
Relu	$f(x)=\begin{cases}0, x\le 0\\ x, x>0\end{cases}$	$f'(x)=\begin{cases}0, x\le 0\\ 1, x>0\end{cases}$

Recurrent neural network

RNN are extensions of traditional feedforward neural networks, capable of handling variable-length sequential inputs, which learn the hidden representation of variable-length input sequences by means of internal recurrent hidden variables, the output of the activation function of the hidden variable at each moment depends on the output of the activation function of the recurrent hidden variable at the previous moment [5]. Fig. 2 shows the structure of RNN, x is the input of the model, $t-1$, t, $t+1$ is the time series, s_t is the memory at time *t* after input x_t, which represents the value of the hidden layer; U is the weight matrix from the input layer to the hidden layer; o is the value of the output layer; V is the weight matrix from the hidden layer to the output layer. The weight matrix W is the value of the hidden layer last time as the weight of this input. f and g are both activation functions:

$$s_t = f\left(Ux_t + Ws_{t-1}\right), \tag{16}$$

$$o_t = g\left(Vs_t\right). \tag{17}$$

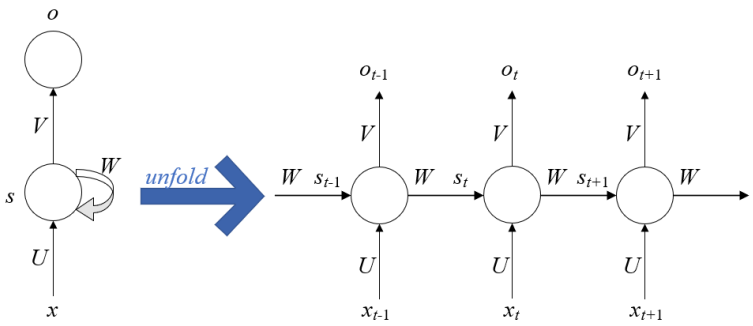


Figure 2. The structure of RNN

The recurrent neural network model is trained by Back Propagation Through Time (BPTT) algorithm, the main use of this algorithm is to process time series, so to be based on time back propagation. The core idea of BPTT algorithm is the same as BP algorithm, is to update the three weight values of W, U and V by gradient descent, so that the error is minimized. The core idea of the BPTT algorithm is the same as that of the BP algorithm. However, RNN can lose gradient after several training sessions, and in some cases, the gradient can be exploded. In response to the difficulty of training recurrent neural networks, various variants of recurrent neural networks have been proposed, such as LSTM, Gated Recurrent Unit (GRU), Bi-LSTM, etc.

Long short-term memory

The main idea of LSTM is to introduce a gate mechanism, so that each LSTM unit can control the degree of historical information retained and remember the current input information, thus capturing potential large-scale sequence dependencies and discarding unimportant features, which are widely used in speech recognition, natural language processing, text compression and handwritten text recognition [6]. A typical LSTM unit is shown in Figure 3.

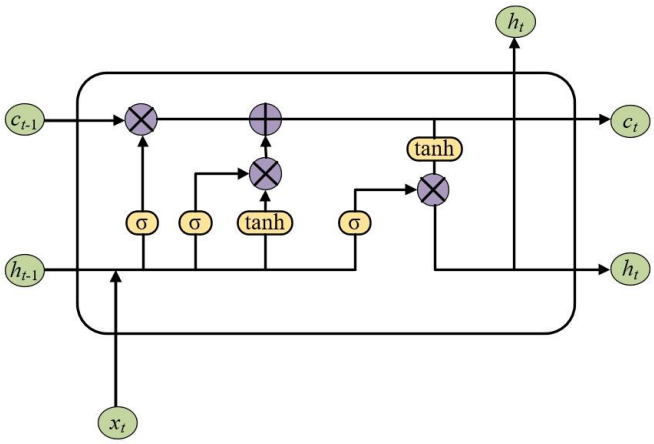


Figure 3. The structure of LSTM

A LSTM unit contains three gates with different functions: an input gate, a forgetting gate and an output gate, which give the LSTM unit the ability to remember information. The input gate determines the proportion of the current state flowing into the current memory cell, the forgetting gate controls the proportion of information in the previous memory cell that is forgotten, and the output gate determines the degree of influence of the current memory cell on the current output.

The information that determines the state of the unit that can be passed. The forgetting gate consists of a sigmoid function that determines the extent to which the unit state can pass based on the output of the previous moment and the input of the current moment. The forgetting gate indicates how much of the state information from the previous time point should be forgotten, which determines how

much of the cell state c_{t-1} from the previous time point is saved to the current cell state c_t. The hidden state h_{t-1} from the previous time point is connected with the input x_t from the current time point to form a new feature quantity, which is then multiplied with the weight parameter W_f, and input to the sigmoid activation function, whose output vector ft is multiplied by the corresponding element of c_{t-1}, $f_t c_{t-1}$ to determine how much of the previous cell state c_{t-1} has been added to the current cell state c_{t-1}. The closer the element of f_t is to 0, the more information in c_{t-1} has been forgotten, and the closer the element of f_t is to 1, the more information in c_{t-1} has been retained. The calculation is shown in equation (3), where W_f and b_f denote the weight parameter and bias term of the sigmoid activation function in the forgetting gate, respectively:

$$f_t = \sigma(W_f \left[h_{t-1}, x_t\right] + b_f). \tag{18}$$

Generate new information that needs to be updated. It consists of two parts, the first step is the input gate to decide which values are used to update, and the second is the tanh function used to generate new candidate values. The input gate indicates how much input information should be remembered at the current point in time, and it determines how much of the input x_t of the cell at the current moment is saved to the cell state c_t. The candidate information c_t at the current moment, determined by the tanh activation function, is determined by multiplying the decision vector it with the corresponding element of the candidate information c_t is $c_t i_t$, to determine how much of c_t is added to the calculation of the unit state c_t. i_t and c_t are shown in equations (4) and (5), respectively. W_i and W_c denote the weight parameters corresponding to sigmoid and tanh in the input gate, respectively, and b_i and b_c denote the corresponding bias terms:

$$i_t = \sigma(W_i \left[h_{t-1}, x_t\right] + b_i), \tag{19}$$

$$c_t = \tanh(W_c \left[h_{t-1}, x_t\right] + b_c). \tag{20}$$

Multiply the previous state value by the coefficient obtained in step 1 to obtain the forgotten part of this unit, and then add the obtained value to the result obtained by multiplying to obtain the new candidate value, which determines the degree of update of each state value:

$$c_t = f_t c_{t-1} + i_t c_t \ . \tag{21}$$

Get the output. An initial output is first obtained by output gating, and then c_t is scaled to between (–1, 1) and then multiply it pairwise with the initial output to obtain the output of the model. The output gate indicates how much of the cell state information should be output at the current point in time, which determines how much of the cell state c_t, at the current moment, is output to the hidden state h_t of the cell. The decision vector o_t of the cell state c_t and the hidden state h_t of the cell, are shown in equations (7) and (8), respectively. W_o and b_o denote the weight parameter and bias term of the sigmoid activation function in the output gate, respectively:

$$o_t = \sigma(W_o \left[h_{t-1}, x_t\right] + b_o), \tag{22}$$

$$h_t = o_t \tanh(c_t). \tag{23}$$

Both the forgetting gate, the input gate and the output gate can be seen to be determined by a value between 0 and 1, which are h_{t-1} and x_t are calculated by the sigmoid function, and although they are in different positions, they measure the necessity of passing the current data backwards: in the forgetting gate, it determines whether c_{t-1} is to be forgotten or not; in the input gate, it determines whether to join or not to join c_t, to update the memory. In the output gate, it determines whether or how important this updated memory is to be input to the next hidden layer.

Conclusion

The development of recurrent neural networks is reviewed and their application background and model characteristics are described in the paper. With the development and fusion of various types of networks, many complex sequence, speech and image problems can be solved, and the variety of recurrent neural networks is quite rich and developing rapidly. With the further deepening of theoretical research and further expansion of application fields, recurrent neural networks will play an increasingly important role.

References

1. Agatonovic Kustrin S., Beresford R. // Journal of pharmaceutical and biomedical analysis. 2000. P. 717–727.
2. Lauriola I., Lavelli A., Aiolli F. // Neurocomputing. 2022. P. 443–456.
3. Jiao M., Wang D., Qiu, J. // Journal of Power Sources. 2020. P. 459.
4. Keren G., Schuller B. // In 2016 International Joint Conference on Neural Networks (IJCNN). 2016. P. 3412–3419.
5. Watson C., Cooper N., Palacio D.N., [et. al] // ACM Transactions on Software Engineering and Methodology (TOSEM). 2022. P. 1–58.
6. Yu Y., Si X., Hu C., [et. al] // Neural computation. 2019. P. 1235–1270.

UDC 621.391

A HYBRID CLASSICATION ALGORITHM BASED ON SVM, ANN AND KNN FOR GESTURE RECOGNITION

Z.M. LIAO

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received February 21, 2023

Abstract. A hybrid model based on SVM, ANN, KNN is proposed, which improves the accuracy compared to the traditional gesture recognition algorithm by combining it with the traditional gesture recognition algorithm.

Keywords: gesture recognition, classification algorithms.

Introduction

Human-computer interaction (HCI) technology has developed rapidly in recent years. There are many kinds of interaction methods, among which, camera-based gesture recognition is mostly used with natural interaction methods. Gesture recognition is used in a wide range of fields, including games, medical care, human-computer interaction, etc. Gesture recognition is a research area of dynamic PC, machine level learning. Gesture recognition, as a typical method of human PC communication, is an area where many experts in academia and industry are working to develop various applications to make association easier and beneficial without wearing any additional devices.

With the development of modern technology and more intelligent lifestyles, people are increasingly longing for a more natural way of interaction. Therefore, it is of great significance to research a natural and comfortable human-computer interaction mode for this phenomenon. At present, several commonly used ways of human-computer interaction are keyboard-based input, mouse-based input, voice, facial expression and gesture recognition technology. The characteristics of natural, intuitive, concise, humanized and flexible gestures can fully stimulate the potential of the hand, without the hand attached to the mouse, keyboard and other external devices, to avoid limitations. And gestures can be reflected in the process of human-computer interaction in a natural and simple way. Gestures have nothing to do with language, and their meanings are common even in different cultures and customs. In addition, there will be communication barriers in both graphic interaction and natural language interaction. Gesture[1] can play a bridging role to solve this problem well. For example, the user does not know the specific use process when using the computer keyboard and mouse, but for how to use and express gestures, it can quickly grasp. Human gestures have diversity and variability, and the efficiency of human-computer interaction can be significantly improved by endowing gestures with specific connotations and inputting them into the computer. Therefore, through the research of gesture recognition, we can better develop a natural and efficient human-computer interaction mode[2], which is in line with the current background needs.

This paper combines the advantages of SVM, ANN and KNN to propose a combination of the three classification algorithms. The small training samples of SVM can overcome the optimization of KNN K-values, ANN improves the running speed by forward propagation, and in addition ANN has a higher accuracy rate. The dataset formed by extracting the eigenvalues of the gesture images is put through a hybrid model, and the respective predictions of SVM, KNN and neural network are voted on before the result is derived.

Gesture segmentation and feature extraction

As images are disturbed and affected by conditions such as light intensity, external environment, differences in equipment performance and various noises (e.g., thermal noise, pretzel noise, etc.) during generation and transmission, the quality of the captured images can be seriously affected, thus affecting the accuracy of gesture recognition. Therefore, before the images are analyzed and processed, they must be pre-processed to improve the quality of the images, enhance the valid information in the images and reduce the useless information in the images to ensure that the results of subsequent image processing are accurate and reliable.

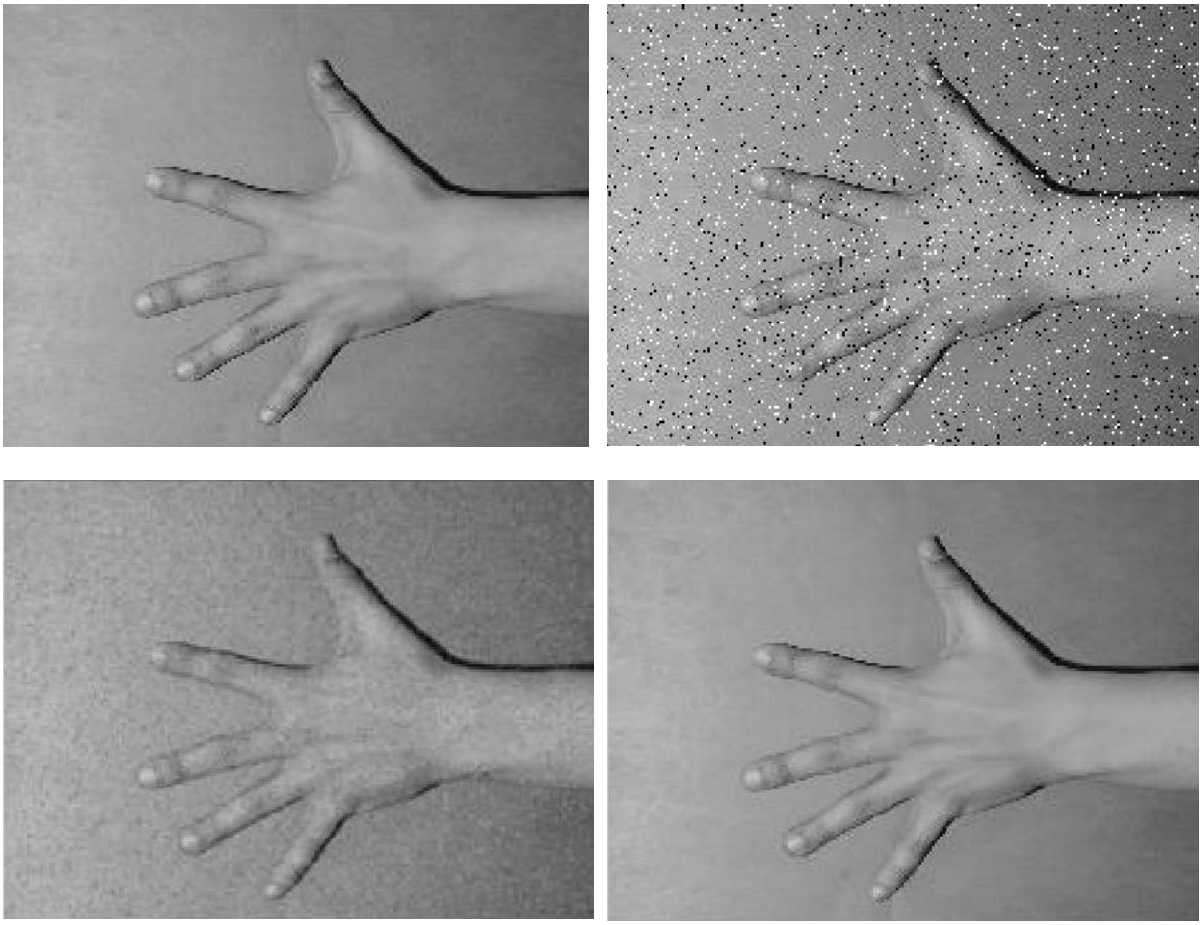


Figure 1. Comparison chart of image smoothing

Before performing gesture recognition, the gesture in the image is segmented and the feature vector of the gesture is extracted. However, due to the influence of external environment and equipment during image acquisition, the input image needs to be processed appropriately before performing gesture segmentation to improve the accuracy of the results. This chapter introduces the common methods about image pre-processing[3-4], such as median filtering and mathematical morphology processing, and gives the results after median filtering and morphology processing.

Theoretical research on traditional classification algorithms

The traditional image recognition process is to first acquire the input image, then detect and segment the desired part of the image (pre-processing), then analyze the features, and finally recognize the image. The content of image pre-processing and image segmentation [5–6] as well as feature extraction have been described earlier. The very important step in the gesture recognition process is the selection of a suitable classification model, and several different classification recognition methods will be presented next.

The KNN algorithm is a traditional machine learning method and a frequently used classical algorithm. It is statistically based, selecting the largest number of samples among the k nearest neighbors

in the experimental sample set and classifying them according to their characteristics. The basic algorithm flow is as follows:

Assume that all sample sets are defined on an N-dimensional space, and usually each sample x can be classified by the feature vector $\{a_1(x), a_2,(x)..., a_r(x)\}$ to represent, where $a_i(x)$ denotes the itch eigenvalue of sample x. Then the similarity between two samples x_i and x_j is generally calculated by using the Euclidean distance, and the formula is shown below:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}. \quad (24)$$

The Euclidean distance formula is used to calculate the distance between two samples, and if the distance value is smaller, it means that the degree of similarity is high and the probability of belonging to the near neighbor between two samples is high, and vice versa, it means that the degree of similarity is low.

Artificial Neural Network (ANN), a mathematical tool abstracted after the brain, is generally simulated as a system consisting of many processing units (i.e., neurons) interconnected with each other. Each of these neurons contains a specific output function, called the excitation function. The connection between each two neurons represents the coefficient weighted value of the signal as it passes through that connection, called the weight. The output of the network is determined by the way the network is connected, the weights and the excitation function. Artificial neural networks are widely used in image classification and recognition systems because of their powerful learning ability. The learning and recognition process of neural networks is the process of updating and adjusting the weights of each neuron.

Support vector machines (SVM) are derived from the theory of statistical methods, which is understood to be based on the exploration of small samples. Therefore, to understand the principles of support vector machines one must be familiar with the theory of statistical methods. SVM has a very intuitive mathematical expression and a clear and explicit mathematical model, and has gained popularity among researchers as soon as it was proposed. This section will focus on the theory, mathematical model and the advantages of the algorithm present in support vector machines.

SVM-ANN-KNN based model for gesture recognition

The previous sections have pre-processed the acquired images, segmented them and extracted the corresponding feature values. The next most important part is to import the extracted feature values into different algorithm models for training and testing to obtain the average accuracy of the final gesture recognition.

Firstly, SVM is used as a classifier for experiments, and the so obtained feature values are loaded into the model, then 10-fold cross validation [7–9] is performed, and the data are divided into 10 parts, one of which is used as the test set and the remaining nine as the training set. The most important part of the SVM model is the selection of parameters, which need to be set as C (penalty factor), the kernel function and the coefficients of the kernel function. The parameter C affects the distance between the support vector and the decision plane, the larger the C, the stricter the classification, there can be no error, the smaller the C, it means that there is greater error tolerance, after many experiments to compare the C set to 350.0, the Gaussian radial basis function is a local strong kernel function, which can map a sample to a higher dimensional space, the kernel function is the most widely used one, whether large The kernel function is one of the most widely used and has better performance for both large and small samples, and it has fewer parameters compared to the polynomial kernel function [10–11], so in most cases when you do not know what kernel function to use, the Gaussian kernel function is used in preference. Finally, the dataset and training set are imported into the SVM model for training and testing, and the accuracy of each round and the final average accuracy are obtained as shown in Table 1.

Table 1. **Comparison of 10-fold cross-validation recognition accuracy of four algorithms**

Accuracy / Model	Test set（%）	Training set（%）
SVM	98,8%	98,3%
ANN	97,3%	97,4%
KNN	96,9%	98,3%
SVM-ANN-KNN	99,3%	99,2%

Conclusion

The algorithm proposed in this paper is verified. First, experiments on gesture segmentation and feature extraction are conducted, and then the extracted feature values are applied to SVM, ANN, KNN and the SVM-ANN-KNN model proposed in this paper. Finally, the four models are compared and tested to compare the gesture recognition accuracy of the four models, and the results show that the accuracy of the SVM-ANN-KNN model proposed in this paper is 99.3%, which is higher than the other three classifiers.

References

1. Karaçalı, B., Ramnath, R., Snyder, W. E. // Pattern Recognition Letters. 2004. Vol. 25. P. 63–71.
2. Kaymak, S., Helwan, A., Uzyn, D. // Procedia computer science. 2017. Vol. 120. P. 126–131.
3. Haralick, R. M., Zhuang, X., Lin, C., Lee, J. S. // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1989. Vol. 37. P. 2067–2090.
4. Lee, J. S. // Computer vision, graphics, and image processing. 1983. Vol. 24. P. 255–369.
5. Yin X, Xie M. // IEEE International Symposium on Computational Intelligence in Robotics and Automation. 2001. P. 438–443.
6. Saravanan, G., Yamuna, G., Nandhini, S. // International Conference on Communication and Signal Processing (ICCSP). 2016. PP. 0462–0466.
7. Tomasi, C., Manduchi, R. // In Sixth international conference on computer vision. 1998. PP. 839-846.
8. Shanableh, T., Assaleh, K., Al-Rousan, M. // IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2007. Vol. 37. P. 641–650.
9. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. // In on The Move to Meaningful Internet Systems. 2003. P. 986–996.
10. Danielsson, P. E. // Computer Graphics and image processing. 1980. Vol. 14. P. 227–248.
11. Imandoust, S. B., Bolandraftar, M. // International journal of engineering research and applications. 2013. Vol. 3. P. 605–610.

UDC 004.62

HIGH DYNAMIC RANGE IMAGE PROCESSING TECHNOLOGY

J.X. FU, J.K. CHEN

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 16, 2023

Abstract. High dynamic range images can be used to capture real environments through digital equipment, and images with high dynamic range are synthesized through specific algorithms in the obtained image information. In the process of image capture, due to the performance of digital equipment, there will be noise in the captured image, which will affect the quality of the image during synthesis. It is necessary to reduce the noise in the image and then synthesize a high dynamic range image. Due to the low dynamic range of traditional display devices, high dynamic range images cannot be directly displayed on traditional devices, and the high dynamic range images are mapped to the dynamic range of the display for display by a method of color scale mapping. Therefore, this paper mainly studies the methods of high dynamic range image synthesis and color scale mapping.

Keywords: high dynamic range image synthesis, multi-exposure image, linear transformation, color scale mapping.

Introduction

The synthesis and display process of high dynamic images involves the principle of camera grayscale image generation and the theory of color scale mapping. This paper briefly analyzes the process of the camera to generate grayscale images and the high dynamic range image synthesis algorithm, and finally introduces the basic theory of color scale mapping.

Digital Image Imaging Process

The imaging process of a digital image can be summarized as that the light passes through the aperture, the shutter, the image sensor, the ADC conversion, and the camera's own color scale mapping curve to generate a displayable grayscale image. In this process, aperture and shutter can be equivalent to a linear response function. Subsequent processing steps: image sensor, ADC conversion, and color scale mapping curve can be equivalent to a camera response function with nonlinear function characteristics.

The most important thing in the synthesis of high dynamic range images is to obtain the nonlinear response function, which describes the brightness values corresponding to different gray scales, so as to restore the original high dynamic dynamic range through the input gray scale image. Among the many solving methods, an effective solution for calibrating the camera response function is to obtain a set of multi-exposure images by shooting the same scene under different exposure time intervals, and then process the grayscale data of each exposure image. Finally, a luminance image that restores the dynamic range in the original scene is synthesized. The core idea of this method is to obtain brightness information of pixels with higher brightness in grayscale images with low exposure, and obtain brightness information of pixels with low brightness in grayscale images with higher exposure. The brightness information is integrated, and the camera response function is fitted according to the properties of the nonlinear response function.

Synthesis of High Dynamic Range Images

The real image in the natural environment has a large dynamic range. From noon to late night, the dynamic range is about 109, while the display device image range of the image is only 102, which is far away. Therefore, the high dynamic range image cannot be completely saved and displayed. In the process of digital image processing, the brightness level is usually divided by 8 bits, and the maximum level range displayed is 0~255. Synthesizing high dynamic range images can not only increase people's perception of real natural scenes and improve the description of image details, but also has very important research significance for subsequent processing processes such as image segmentation and edge detection [1–2].

Usually, images are acquired by converting photon signals into image pixel information through the CCD (charged couple device) sensor in the camera [3]. Developers develop the latest CCD components to improve the dynamic range of image sensor components. Mitsunaga proposed to design an image sensor to change the exposure by changing the spatial position, and then, by placing an optical mask in front of the photosensitive array of the image sensor, because the transmittance of the incident light corresponding to each position is different, the image array Different exposures can be obtained at different positions, and finally the image sensor will form a frame of high dynamic range image. Others have proposed similar design schemes. For example, Tublin proposed to design a camera that uses the principle of the photosensitive unit of the camera to first detect the difference in incident light between adjacent pixels, then quantify the size of the incident light, and finally make the camera generate a high dynamic range image [3]. The above two design ideas are basically to improve and design the lighting of the camera image sensor, and the ultimate goal is to improve the adaptability of the camera to the dynamic range, but this will increase the development cost of the camera to a certain extent The effect of a large dynamic range is not very good, and the cost of the hardware itself is relatively high, which limits the further development and application of the camera, resulting in the development of a high dynamic camera that cannot be widely used by mass consumers.

The basic principle of the software synthesis algorithm is to use ordinary camera equipment to obtain image sequences under different exposure conditions in the same scene for the same scene, and then use various algorithms to synthesize a high dynamic range algorithm. The details in the image are composited into one image to expect greater dynamic range. The software algorithm has a certain effect on images in static scenes, but if the object is displaced during the image sequence synthesis process, it will cause ghosting, that is, object ghosting may occur. This will result in poor quality of the generated high dynamic image, and even affect the visual effect.

According to the above, the key to synthesizing high dynamic range images is to find the nonlinear response function g, that is, to find the brightness values corresponding to different gray scales, so as to restore the dynamic range of the input image. An effective way to calibrate the camera response function is to capture a set of multi-exposure images by shooting the same scene under a set of different exposure time conditions, and then extract, organize, and analyze the grayscale data of each exposure image to synthesize A luminance image that represents the light distribution of the original scene. The core idea of this method is to extract the brightness information of pixels with higher brightness in the grayscale image with lower exposure, and extract the brightness information of pixels with lower brightness from the grayscale image with higher exposure. The brightness information is integrated, and the camera response function is fitted according to the properties of the nonlinear response function. Debevec's algorithm and Mitsunaga's algorithm are representative of the methods for finding the fitted nonlinear response function.

Color Scale Mapping

Level mapping is a general term for a method that "converts" a high dynamic range image into a viewable image. The purpose is to solve a common problem that plagues people when viewing high dynamic range images, that is, CRT, LCD, printers and other images. The displayed dynamic range is limited.

In image transformation, how to maintain image details, colors, and other important information for expressing the original scene requires a large contrast attenuation to transform the scene brightness to a displayable range. Tone mapping is essentially to solve this kind of problem. Tone mapping is a

computer graphics technique for approximating high dynamic range images on limited dynamic range media [4]. CRT, LCD monitors or printouts, projectors, etc. all have limited dynamic range. Tone mapping is inherently expressive and its goals vary from application to application. Different occasions have different requirements for images. In some occasions, some require high-quality images, some require more details in the image, and some require the maximization of image contrast. In practical applications, the integrity of the display device in the display brightness range may not be complete, but people require that the real scene and the display image match.

The values of all pixels in the HDR image are directly related to the measured luminosity values of the actual environment, and the stored pixel values are linear. This property is called scene correlation. The color space of the current LDR image is for a specific output device, and their output is related. The data displayed in this way is low dynamic range data, which meets the range of the output device, but there will be recorded scenes that do not correspond to the color space. Therefore, the color values contained in the LDR image are limited, and the information describing the real scene can be as close as possible to the original data of the scene. HDR images can only be displayed on high dynamic range devices. A specific method is required to map the pixel values of high dynamic range images to the color space of the device. This mapping method is called tone mapping. The general model of color scale mapping is shown in Figure 1.

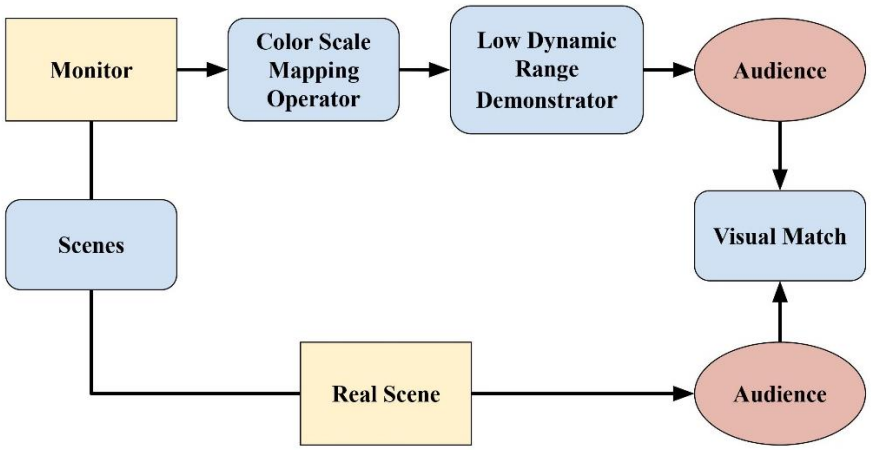


Figure 1. General model for color scale mapping

The tone mapping algorithm has been diversified in recent years. The tone mapping filter is its typical representative, $L = Y / Y + 1$, the filter is the radianc value, and the mapping of Y in the [0, ∞] domain, to the range [0, 10] show the output. Algorithms based on the gradient domain are more complex. This type of algorithm is concerned with maintaining contrast without considering the mapping of brightness. The contrast or the brightness ratio of different brightness regions is the most attractive source of inspiration for this type of method. This tone mapping preserves better contrast detail and typically produces very sharp images, but this tone mapping results in a flattened overall image contrast. The most typical examples of tone mapping methods are: high dynamic range image perception framework and gradient domain high dynamic range compression. The implementation method of high dynamic range image tone mapping is an idea obtained from anchoring theory of lightness perception. Lightness constancy and its spectacular failures in the human visual system can be fully understood in this theory. The central theory of this tone mapping method is to map high dynamic range images into illuminated uniform structures or regions and to compute local luminance values. Merge according to the luminance ratio of all unused areas, so as to calculate the pure luminance value of the image. The key is to correspond the constant brightness value to the brightness of the brightness positioning lighting, or which brightness value is perceived as white in the scene. This kind of tone mapping retains the natural color of high dynamic images, because its implementation method does not affect the local contrast. The tone mapping algorithm can be roughly divided into two categories according to the spatial correlation: local tone algorithm (spatial correlation) and Global tone algorithm (space independent). Because most tone mapping algorithms are designed for the visual reproduction of tone levels of grayscale high dynamic range images, they do not take the brightness range of the image into account [4]. After compression, the color gamut of the image is compressed at the same time. Will cause

the image color to shift visually. The research on the tone mapping algorithm in the field of image processing is a hot spot. The tone mapping algorithm can map the brightness range of the real world scene to the range that the output device can output. The characteristics of the human visual system and the compressed luminance range are also considered while preserving the perceptual quality of the image. So far we have not been able to create an objective quality criterion from quantitative data. All tone mapping algorithms will lose information in the process of compressing the brightness range, such as contrast, brightness, image details, etc., and this information are only retained in the image in a targeted manner.

Conclusion

This paper mainly introduces the relevant theoretical basis of high dynamic range processing technology, and provides theoretical background knowledge for future algorithms. First, the digital image imaging process is introduced, and the response process of digital equipment is explained; secondly, the theory of high dynamic range image synthesis is introduced, and the synthesis method of high dynamic range images is explained in two ways: software and hardware. Finally, the color scale mapping of high dynamic range images is introduced, and the global color scale mapping algorithm and the local color scale mapping algorithm are briefly introduced.

References

1. Lin K.Y., Wu J.H., Xu L.H. // Chinese Journal of Image and Graphics. 2005. Vol. 10. P. 1–10.
2. Duan R.L., Li Q.X., Li Y.H. // Optical Technology. 2005. Vol. 31. P. 415–419.
3. Liu W.H. // Electronic World. 2014. P. 178–179.
4. Wan X.X., Xie D.H. // China Printing and Packaging Research. 2009. P. 1–6.

УДК 654.9

ОПТИМИЗАЦИЯ ПРОЦЕССА МОНИТОРИНГА НЕИСПРАВНОСТЕЙ В КОММУТАТОРАХ СИСТЕМ ОПОВЕЩЕНИЯ

А.П. ТУРЛАЙ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2023

Аннотация. Рассмотрен коммутатор производства компании СЕНСОР-М. Освещен основной функционал данного коммутатора, выявлены особенности в работе алгоритма анализа и принятия решения о состоянии линий связи. Приведены результаты исследований, определены критерии влияющие на возможность расширения функционала (увеличения количества линий связи). Определены методы, позволяющие, не прибегая к глобальному расширению аппаратной составляющей увеличить количество коммутируемых каналов, не ухудшая скоростных характеристик работы каналов.

Ключевые слова: система оповещения, блок коммутации и контроля, оптимизация, линия связи, АРМ оператора, статус линий, живучесть комплекса.

Введение

Современные системы оповещения используют сетевую архитектура программно-конфигурируемых сетей на основе контролера и коммутаторов. К системам оповещения в настоящее время применяются жесткие требования как к их функционалу, так и к живучести.

Для исключения случаев несвоевременного оповещения персонала либо населения в любой системе оповещения предусмотрены меры, обеспечивающие ее бесперебойную работу, такие как резервирование оборудования, резервирование линий связи и т.д. Один из путей повышения эффективности работы таких систем связан с оптимизацией работы блоков коммутации и контроля (БКК).

Целью данных исследований является оптимизация работы БКК в части определения статуса коммутируемых линий на примере системы компании СЕНСОР-М.

Предметом исследования определим временные затраты БКК, в частности его интеллектуального модуля на получение статуса одной линии связи. Для этого необходимо решить следующие задачи:

– получить общее время получения статуса;
– получить данные о каждой составляющей машинного времени микроконтроллера и ПК;
– провести анализ полученных данных, выяснить максимальные временные затраты;
– предложить варианты уменьшения временных затрат.

Блок коммутации и контроля

Локально БКК в системе оповещения на базе комплекса производственной громкоговорящей командно-поисковой связи, оповещения и управления эвакуацией расположен в стойке, рядом с компьютером [1].

На рис. 1 изображена схема системы оповещения, ее составные части и место БКК в общей структуре системы. Устройство БКК во время сеанса оповещения выполняет функции коммутатора выходов радиотрансляционных усилителей мощности, с контролем нагрузки в линии трансляции (радиофидере) и локализацией критических ошибок в них (короткое замыкание, обрыв).

Основным элементом БКК является «Интеллектуальный модуль» он принимает команды от ПК и управляет всеми внутренними модулями устройства. По запросу модуль коммутации передает информацию от измерительных модулей в интеллектуальный модуль для анализа и формирования статуса линий связи. Все статусы передаются на интерфейс компьютера оперативного дежурного для контроля состояния устройств и линий связи [2]. Структурная схема БКК представлена на рис. 2.

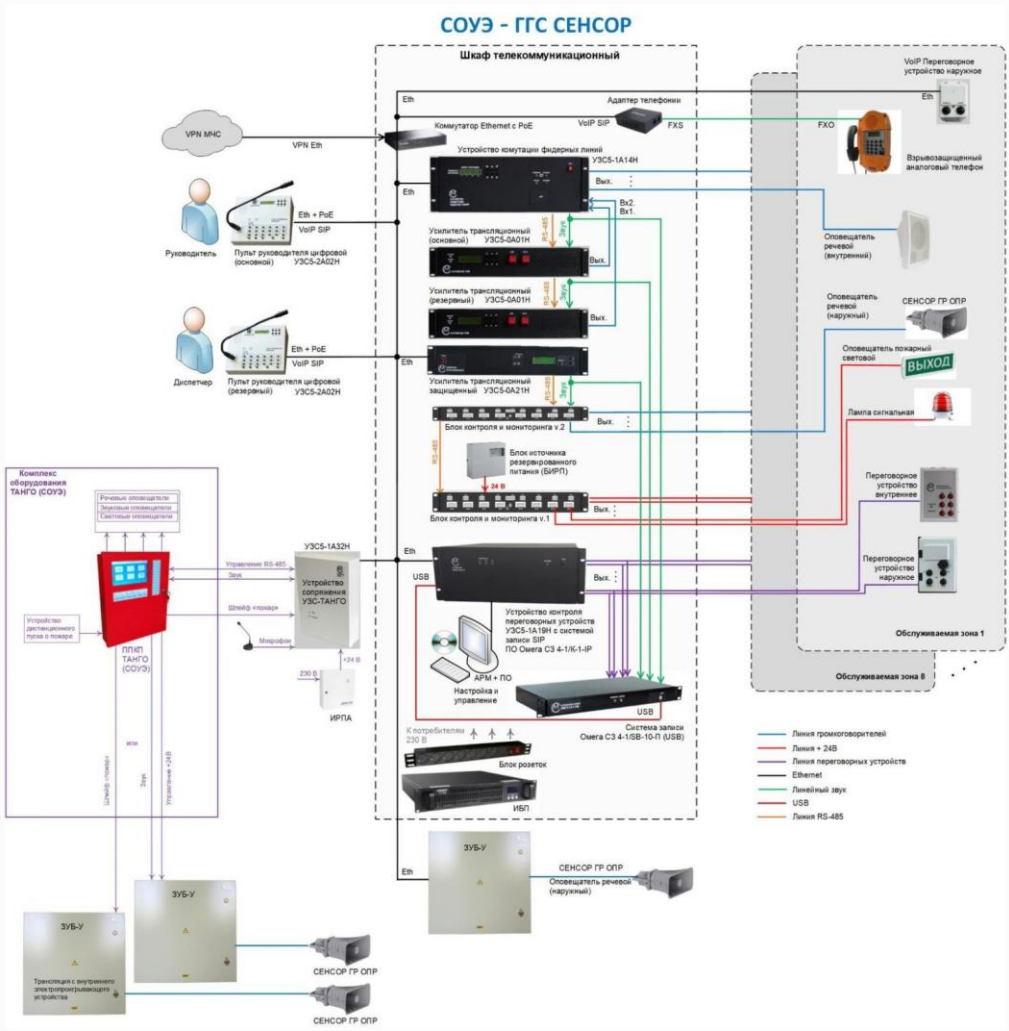


Рис. 1. Схема системы оповещения

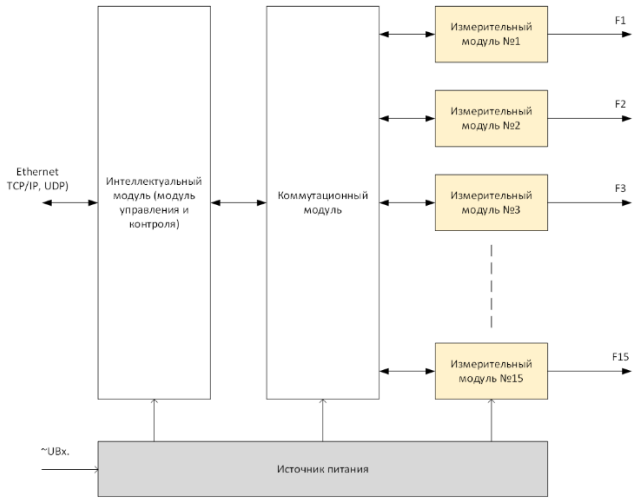


Рис. 2. Структурная схема БКК

Обеспечение живучести

Одной из важных задач устройства является поддержание живучести комплекса оповещения. Для оперативного реагирования на различные рода аварийные ситуации и быстрого принятия решения по их устранению в каналах оповещения производиться постоянный мониторинг состояний как самого устройства БКК, так и линий связи. На практике задачу сбора данных о состоянии устройств и линий оповещения выполняет интеллектуальный модуль, входящий в состав БКК. Этот модуль с заданной периодичностью проводит мониторинг состояния каналов связи и определяет их состояния:

– норма;

– обрыв канала оповещения;

– короткое замыкание канала оповещения.

Информацию о состоянии линий связи предоставляют датчики интегрированные в линию связи. Обработку полученных от датчиков данных, принятие решения о состоянии линии связи осуществляет контроллер со встроенным программным обеспечением входящий в состав интеллектуального модуля. Информация о состоянии линии поступает от БКК по каналу Ethernet на АРМ оператора, сохраняется и отображения в интерфейсе оперативного дежурного [3].

Алгоритм работы контроллера учитывает множество параметров и их значения для оценки состояния каналов связи. На опрос состояний линий связи тратиться достаточно большой машинный ресурс. Контроллер постоянно отслеживает время сканирования, активирует внутренние периферийные устройства, переключает каналы телеметрии, производит обработку поступившей информации, устанавливает статус линии связи и передавать пул статусов на АРМ оператора. Временные затраты алгоритма показаны в табл. 1.

Табл. 1. **Временные затраты алгоритма**

Общее время работы МК				Общее время работы ПК	
Время активации периферии (t_1), нс	Время считывания показаний с датчиков (t_2), нс	Время анализа показаний датчиков (принятие решения о состоянии линии связи) (t_3), нс	Время передачи статуса на ПК (t_4), нс	Время приема и сохранения данных в БД, (t_5), нс	Время отображения на АРМ ОД (t_6), нс
10	60	1000	300	150	100

Временная нагрузка на контроллер приводит к ограничениям по числу линий связи, не возможности расширения линий оповещения без установки дополнительного оборудования, что в свою очередь приводит к увеличению стоимости системы.

Учитывая все это сформируем критерии оптимизации системы (комплекса) оповещения:

– скорость принятия решения;

– число линий связи;

– стоимость оборудования коммутации.

Анализ временных затрат

Общее время на принятие решения будет равно сумме всех времен из табл. 1:

$$t_{\text{общ}} = t_1 + t_2 + \ldots + t_6 = 1620.$$

Общее время на обработку одного канала связи примерно будет составлять:

$$t_{\text{мк}} \approx t_1 + t_2 + t_3 + t_6 = 10 + 60 + 1000 + 300 = 1370 \text{ нс}.$$

Как видно из табл. 1 t_3 – анализ показаний датчиков занимает больше всего времени. В основном это зависит от характеристик контроллера и особенностей алгоритма работы.

В условиях заданных ограничений на быстродействие и память, одно из направлений оптимизации процесса принятия решения о неисправности линий связи, связано с уменьшением $t_{\text{мк}}$.

Одним из вариантов решения данной задачи может быть перенос обязанностей с контроллера на более производительные узлы системы оповещения, такие как компьютер оперативного управления.

В табл. 2 приведены данные после переноса части функционала с контроллера на компьютер управления.

Табл. 2. **Временные затраты алгоритма после переноса части функционала**

Общее время работы МК			Общее время работы ПК		
Время активации периферии (t_1), нс	Время считывания показаний с датчиков (t_2), нс	Время передачи статуса на ПК (t_4), нс	Время анализа показаний датчиков (принятие решения о состоянии линии связи) (t_3), нс	Время приема и сохранения данных в БД, (t_5), нс	Время отображения на АРМ ОД (t_6), нс
10	60	300	60	150	100

Из табл. 2 видно, что общее время контроллера уменьшилось. Общее время на принятие решения также уменьшилось более чем в 2 раза:

$$t_{\text{мк}} \approx t_1 + t_2 + t_3 + t_6 = 10 + 60 + 300 = 370 \text{ нс},$$

$$t_{\text{общ}} = t_1 + t_2 + \ldots + t_6 = 680.$$

Это решение позволяет разгрузить контроллер, возложив на него дополнительный функционал увеличив количество каналов оповещения, не прибегая к применению дополнительного коммутационного оборудования, что повышает эффективность работы системы.

Вывод

Были проведены исследования, направленные на выявление слабых мест в алгоритме обработки данных и принятии решения о статусе линий связи в системах оповещения использующих БКК. Выявлено, что время на обработку данных параметров линий связи на встроенном в интеллектуальный модуль контроллера БКК нерационально т.к. характеристики и нагрузочная способность МК ограничены в текущем исполнении модуля, что видно из приведенных выше табличных данных. Эти ограничения не позволяют расширить количество каналов коммутации, не прибегая к разрастанию архитектуры системы, что приводит к увеличению ее стоимости. Таким образом оптимизация процесса получения статуса линий связи будет актуальна при переносе алгоритма анализа данных на верхний уровень архитектуры системы. Это уменьшает время обработки и принятия решения, высвободив тем самым ресурс контроллера для увеличения количества каналов связи.

OPTIMIZATION OF THE PROCESS OF FAULT MONITORING IN THE SWITCHES OF WARNING SYSTEMS

A.P. TURLAI

Abstract. The switchboard manufactured by SENSOR-M is considered. The main functionality of this switch is highlighted, the features of the algorithm of analysis and decision-making on the state of communication lines are revealed. The results of the research are presented, the criteria affecting the possibility of expanding the functionality (increasing the number of communication lines) are determined. Methods have been identified that allow, without resorting to a global expansion of the hardware component, to increase the number of switched channels without compromising the speed characteristics of the channels.

Keywords: notification system, switching and control unit, optimization, communication line, operator's ARM, line status, survivability of the complex.

Список литературы

1. ООО «СЕНСОР-М». Руководство по эксплуатации БКК. 2022.
2. ООО «СЕНСОР-М». [Электронный ресурс]. URL: https://sensor-m.com.
3. Texas Instruments. Datasheet DSP. 2010.

СВЕДЕНИЯ ОБ АВТОРАХ

1.	Богуш Рихард Петрович	– д.т.н., заведующий кафедрой вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой
2.	Борисенко Олег Федерович	– к.ф-м.н., доцент кафедры вышей математики, БГУИР
3.	Борискевич Анатолий Антонович	– д.т.н., профессор, профессор кафедры инфокоммуникационных технологий БГУИР
4.	Борискевич Илья Анатольевич	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
5.	Бысов Анатолий Александрович	– к.т.н., доцент, начальник цикла кафедры связи Военной академии Республики Беларусь
6.	Ван Ин	– магистрант кафедры инфокоммуникационных технологий БГУИР
7.	Вань Цзывэй	– магистрант кафедры инфокоммуникационных технологий БГУИР
8.	Вишняков Владимир Анатольевич	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
9.	Врублевский Сергей Сергеевич	– адъюнкт кафедры связи Военной академии Республики Беларусь
10.	Вэй Цзыцзянь	– магистрант кафедры инфокоммуникационных технологий БГУИР
11.	Вэй Шиши	– магистрант кафедры инфокоммуникационных технологий БГУИР
12.	Гавриленко Валерия Сергеевна	– преподаватель УО БГУИР филиал МРК
13.	Гао Хаосюань	– магистрант кафедры инфокоммуникационных технологий БГУИР
14.	Гао Ялу	– магистрант кафедры инфокоммуникационных технологий БГУИР

15.	Голубенок Аким Александрович	– студент специальности «Вычислительные машины, системы и сети» Полоцкого государственного университета имени Евфросинии Полоцкой
16.	Громов Владимир Анатольевич	– магистрант кафедры инфокоммуникационных технологий БГУИР
17.	Жэнь Сюнь Хуань	– аспирант кафедры инфокоммуникационных технологий БГУИР
18.	Игнатьева Светлана Александровна	– аспирант кафедры вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой
19.	Касанин Сергей Николаевич	– к.т.н., доцент, заместитель директора по науке научно-производственного республиканского унитарного предприятия «Объединенный институт проблем информатики»
20.	Конопелько Валерий Константинович	– д.т.н., доктор технических наук, почетный профессор БГУИР
21.	Кучеров Сергей Владимирович	– магистрант кафедры инфокоммуникационных технологий БГУИР
22.	Лапицкая Василина Александровна	– к.т.н., научный сотрудник лаборатории нанопроцессов и технологий Института тепло- и массообмена имени А.В.Лыкова НАН Беларуси
23.	Левоненко Иван Игоревич	– студент факультета информационной безопасности БГУИР
24.	Ли Бои	– студент факультета информационных технологий и управления БГУИР
25.	Ловецкий Михаил Юрьевич	– аспирант кафедры инфокоммуникационных технологий БГУИР, младший научный сотрудник лаборатории нанопроцессов и технологий Института тепло- и массообмена имени А.В. Лыкова НАН Беларуси
26.	Лю Фаньюй	– магистрант кафедры инфокоммуникационных технологий БГУИР
27.	Ляо Чжунминь	– аспирант кафедры инфокоммуникационных технологий БГУИР

28.	Ма Цзюнь	– аспирант кафедры инфокоммуникационных технологий БГУИР
29.	Матюшенко Антон Дмитриевич	– учащийся УО «Национальный детский технопарк»
30.	Мацкевич Вадим Владимирович	– аспирант кафедры информационных систем управления ФПМИ БГУ
31.	Михно Кирилл Валерьевич	– курсант военного факультета БГУИР
32.	Нгуен Ван Бась	– магистрант кафедры кафедры инженерной и компьютерной графики БГУИР
33.	Петров Сергей Николаевич	– доцент кафедры защиты информации БГУИР
34.	Полуян Татьяна Владимировна	– аспирант кафедры инфокоммуникационных технологий БГУИР
35.	Протько Мария Андреевна	– студент кафедры электронных вычислительных машин БГУИР
36.	Рощупкин Яков Викторович	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
37.	Руденя Даниил Александрович	– учащийся УО «Национальный детский технопарк»
38.	Рудиков Станислав Игоревич	– магистр технических наук, заместитель директора по информационным технологиям Унитарного предприятия «НТЦ «ЛЭМТ» БелОМО»
39.	Саломатин Сергей Борисович	– к.т.н., доцент., доцент кафедры инфокоммуникационных технологий БГУИР
40.	Сидоренко Сергей Александрович	– магистрант кафедры инфокоммуникационных технологий БГУИР
41.	Сюй Вэйсюань	– студент факультета информационных технологий и управления БГУИР
42.	Томашевич Никита Александрович	– студент специальности «Вычислительные машины, системы и сети» Полоцкого государственного университета имени Евфросинии Полоцкой

43.	Турлай Андрей Петрович	– аспирант кафедры инфокоммуникационных технологий БГУИР
44.	У Хаожань	– магистрант кафедры инфокоммуникационных технологий БГУИР
45.	Усевич Андрей Владимирович	– магистрант кафедры инфокоммуникационных технологий БГУИР
46.	Фу Цзяньсян	– магистрант кафедры инфокоммуникационных технологий БГУИР
47.	Хаджинова Наталья Владимировна	– старший преподаватель кафедры информационных технологий автоматизированных систем БГУИР
48.	Харченко Александр Владимирович	– учащийся УО БГУИР филиал МРК
49.	Хе Тао	– магистрант кафедры инфокоммуникационных технологий БГУИР
50.	Хэ Жуньхай	– студент факультета информационных технологий и управления БГУИР
51.	Цветков Виктор Юрьевич	– д.т.н., профессор, заведующий кафедрой инфокоммуникационных технологий БГУИР
52.	Цзэн Пэн	– магистрант кафедры инфокоммуникационных технологий БГУИР
53.	Чжан Бовэнь	– студент факультета информационных технологий и управления БГУИР
54.	Чжан Жунлян	– студент факультета информационных технологий и управления БГУИР
55.	Чжан Хэнжуй	– студент факультета информационных технологий и управления БГУИР
56.	Чжао Иань	– магистрант кафедры инфокоммуникационных технологий БГУИР
57.	Чжоу Цюаньхуа	– студент факультета информационных технологий и управления БГУИР
58.	Чжун У	– студент факультета информационных технологий и управления БГУИР

59.	Чижик Сергей Антонович	– д.т.н., профессор, академик НАН Беларуси, главный научный сотрудник лаборатории нанопроцессов и технологий Института тепло- и массообмена имени А.В. Лыкова НАН Беларуси
60.	Чэнь Имин	– магистрант кафедры инфокоммуникационных технологий БГУИР
61.	Чэнь Цзикэ	– магистрант кафедры инфокоммуникационных технологий БГУИР
62.	Чэнь Чжэин	– магистрант кафедры инфокоммуникационных технологий БГУИР
63.	Шевчук Оксана Геннадьевна	– кандидат технических наук, доцент кафедры инфокоммуникационных технологий БГУИР
64.	Шкадаревич Алексей Петрович	– д.ф-м.н., профессор, академик НАН РБ, директор Унитарного предприятия «НТЦ «ЛЭМТ» БелОМО»
65.	Ян Цзысяо	– магистрант кафедры инфокоммуникационных технологий БГУИР

no