Министерство образования Республики Беларусь

Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

КОДИРОВАНИЕ И ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ В ИНФОКОММУНИКАЦИЯХ

МАТЕРИАЛЫ МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ
(Минск, 4 апреля 2019 г.)

Минск, 2019

УДК 621.391.7:654
ББК 32.811.4+32.88
К57

К57 **Кодирование** и цифровая обработка сигналов в инфокоммуникациях: материалы междунар. науч.-практ. конф. (Республика Беларусь, Минск, 4 апреля 2019 года) / редкол.: В. К. Конопелько, В. Ю. Цветков, Л. А. Шичко – Минск : БГУИР, 2019. – 136 с.: ил.

ISBN 978-985-543-496-3.

Сборник содержит статьи, тематика которых посвящена научно-теоретическим и практическим разработкам в области цифровой обработки сигналов, теории кодирования, обработки и передачи изображений, защиты инфокоммуникационных систем и сетей.

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

Научное издание

Корректор *О. В. Бойправ*
Ответственный за выпуск *В. К. Конопелько*
Компьютерный дизайн и верстка *В. В. Чепикова*

Подписано в печать 01.04.2019. Формат 60×84 ⅛. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 16,28. Уч.-изд. л. 13,3. Тираж 60 экз. Заказ 74.

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

ISBN 978-985-543-496-3

СОДЕРЖАНИЕ

CONTENTS

СЕКЦИЯ 1
ТЕОРИЯ КОДИРОВАНИЯ И ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

UDC 621.391.14

INTERLEAVING BASED ON LINEAR BLOCK CODE

REN XUNHUAN, O.J.M. Al-FURAIJI, V.K. KONOPELKO

Belarusian state university of informatics and radioelectronics, Republic of Belarus

Submitted 19 March 2019

Abstract. In this paper the methods of transforming one-dimensional linear code into two-dimensional symmetric and asymmetric recursive scattering information code are analyzed. The method has transformed the encoded information sequence into a discrete sequence to combat the burst noisy environment. Channel coding is effective when detecting and correcting single errors are not too long strings and is ineffective for long burst errors caused by interference and fading of the mobile communication channel. However, the interleaving technique can solve the above problem without adding extra redundancy.

Keywords: symmetric and asymmetric method, interleaving technology, channel coding, burst interference.

Introduction

In recent years, there has been an increasing demand for efficient and reliable digital data transmission and storage systems. This demand has been accelerated by the emergence of large-scale, high-speed data networks for the exchange, processing, and storage of digital. A major concern of the designer is the control of errors so that reliable reproduction of data can be obtained.

In digital communication, due to inherent noise characteristics and fading characteristics, there are always different levels of interference and fading during information transmission, resulting in errors in signal transmission. In order to reduce the information error rate and improve the reliability of information transmission, channel coding techniques such as Hamming code, cyclic code and convolutional code for error correction have been invented. However, with the development of technology and the increasing demand for information accuracy, the limitations of these technologies have become increasingly prominent. For example, channel coding is only effective when detecting and correcting single errors and not too long error strings and generates multiple consecutive errors when the channel generates burst interference. Such Hamming codes, cyclic codes, and convolutional codes are not enough.

In this regard, the interleaving technology have been innovated and invented. As a new coding technology, interleaving coding technology is mainly used for memory channels, especially in wireless channels, to correct some bit errors and some burst errors.

A data interlace is often added to the end of the transmitter and a deinterlace is located behind the receiver, and the interleaved code is used to actively modify the channel, and a memorized burst channel is transformed into an independent memoryless channel through the interleaving and deinterleaving channel. Thereby the burst errors of the channel are spread out.

In this way, if some burst errors are occurred, thanks to the existence of the interleaving module, the errors are divided to some individual ones and then it is possible to correct these separated errors and revive the original information by use of the channel coding error correction ability [1].

Interleaving coding principle

The basic structure of the system is presented in Fig. 1.

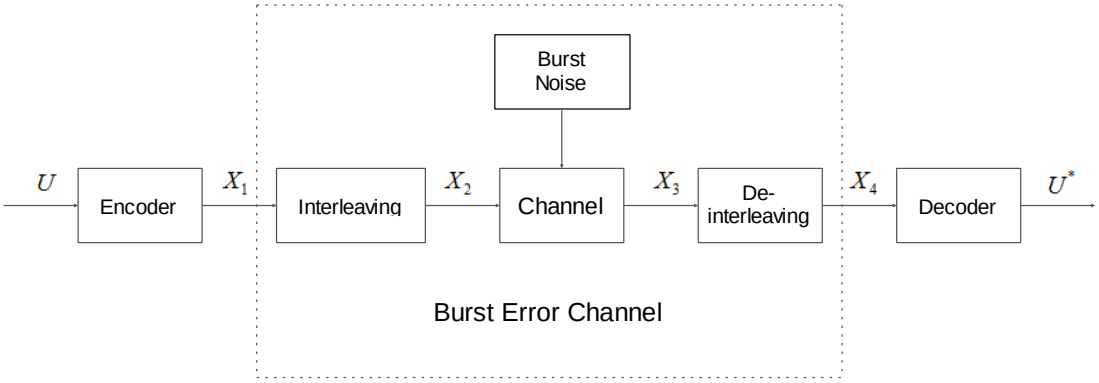


Fig. 1. Interleaved coding principle block diagram

The transmitted information can be represented by symbols. The function of the linear encoder is to encode the message which consists of binary digits string. The main task of the communications system is to convey the information from one point to another without distortion. But the ideal channel without any noise is not exist, the information always exposed to the noise which causes degradation of the signal. The detailed behavior of communication systems in the presence of noise is a lengthy study. The authors are concerned mostly with the interleave and deinterleave block, transmitting the message to the data interleave, adding a de-interleave to the receiving end, using the interleaving method to actively modify the encoded message, and transforming a memorized burst channel into an independent memoryless channel through the interleaving and deinterleaving channel, thereby the burst errors of the channel are spread out. The generation of interleave greatly improves the reliability of information transmission. The channel error correction code is to adapt to the channel, by adding redundancy to high transmission reliability, only suitable for anti-random interference. Different interleaving codes are used for error correction. They transform the channel without adding redundancy, suitable for anti-burst interference [2].

The common methods of the interleaving are packet interleaving, convolutional interleaving and random interleaving. The process of the analysis will conduct with the help of the matrix. The information which needed to transmit presented like following:

$$X_1 = (x_0, x_1, x_2, ..., x_{14}, x_{15}). \quad (1)$$

The interleaving memory is written by the column and read by the row.

$$X_1' = \begin{bmatrix} x_0 & x_4 & x_8 & x_{12} \\ x_1 & x_5 & x_9 & x_{13} \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{bmatrix} \quad (2)$$

The output information of the interleaving block which will sent to the channel:

$$X_2 = (x_0, x_4, x_8, x_{12}, x_1, x_5, x_9, x_{13}, x_2, x_6, x_{10}, x_{14}, x_3, x_7, x_{11}, x_{15}). \quad (3)$$

In case, when two burst errors appear, one of them is started from the x_0 and ended in the x_{12} which means there are 4 successive errors, and another one is in the 2 continual position, from x_9 to x_{13}, the error is donated as:

$$X_3 = (x_0', x_4', x_8', x_{12}', x_1, x_5, x, x_{13}, x_2, x_6, x_{10}, x_{14}, x_3, x_7, x_{11}, x_{15}). \quad (4)$$

In the receiver information will saved in another storage after leaving the interleaving modular which is written by row and read by column:

$$X_3' = \begin{bmatrix} x_0' & x_4' & x_8' & x_{12}' \\ x_1 & x_5 & x_9' & x_{13}' \\ x_2 & x_6 & x_{10} & x_{14} \\ x_3 & x_7 & x_{11} & x_{15} \end{bmatrix}. \quad (5)$$

The output of the deinterleaving modular is:

$$X_4 = (x_0', x_1, x_2, x_3, x_4', x_5, x_6, x_7, x_8', x_9', x_{10}, x_{11}, x_{12}', x_{13}', x_{14}, x_{15}). \quad (6)$$

It is obvious that with the help of the interleaving modular, two burst errors, with number of the continual error are 4 and 5 respectively, will change to the random independent error.

Introduction to symmetric and asymmetric interleaving [3,4]

Based on the 2×2 matrics, design the $2^N \times 2^N$ (N-integer) matrics, the interleaved matrices are as presented in Fig. 2.

0	4	8	12
1	5	9	13
2	6	10	14
3	7	11	15

a

0	8	2	10
12	4	14	6
3	11	1	9
15	7	13	5

b

0	4	6	10
8	12	14	2
15	3	9	13
7	11	1	5

c

0	3	6	8
10	12	14	1
4	7	9	11
13	15	2	5

d

Fig. 2. Symmetric and Asymmetry matrices 4×4

The permutation functions on {0,1,2,3} as follows [3]:

$$\begin{aligned}
&h_{6i+1}(0)=0, \quad h_{6i+1}(1)=1, \\
&h_{6i+2}(0)=0, \quad h_{6i+2}(1)=2, \\
&h_{6i+3}(0)=0, \quad h_{6i+3}(1)=2, \\
&h_{6i+4}(0)=0, \quad h_{6i+4}(1)=1, \quad h_{6i+1}(2)=2, \quad h_{6i+1}(3)=3; \\
&h_{6i+5}(0)=0, \quad h_{6i+5}(1)=3, \quad h_{6i+2}(2)=1, \quad h_{6i+2}(3)=3; \\
&h_{6i+6}(0)=0, \quad h_{6i+6}(1)=3, \quad h_{6i+3}(2)=3, \quad h_{6i+3}(3)=1; \\
&\qquad\qquad\qquad\quad h_{6i+4}(2)=3, \quad h_{6i+4}(3)=2; \\
&\qquad\qquad\qquad\quad h_{6i+5}(2)=1, \quad h_{6i+5}(3)=2; \\
&\qquad\qquad\qquad\quad h_{6i+6}(2)=2, \quad h_{6i+6}(3)=1;
\end{aligned} \quad (7)$$

and $h_i(4k+m)=h_i(m)$ for all nonnegative integers i, k and m.

$$\begin{aligned} &g_{6i+1}(n) = h_5(n), \quad g_{6i+2}(n) = h_3(n), \quad g_{6i+3}(n) = h_1(n); \\ &g_{6i+4}(n) = h_6(n), \quad g_{6i+5}(n) = h_2(n), \quad g_{6i+6}(n) = h_4(n); \end{aligned} \tag{8}$$

for all nonnegative integers i.

$$f(n,a) = [g_{a+1}(n) + \sum_{b=1}^{a} h_{a-b+1}([\frac{n}{4^b}])](\mathrm{mod}\,4). \tag{9}$$

Where $[\bullet]$ is the floor function ($[x]$ is the greatest integer less than or equal to x), and define

$$\begin{aligned} r(n,a) &= \begin{Bmatrix} 0, \text{ if } f(n,a) \in \{0,1\} \\ 1, \text{ if } f(n,a) \in \{2,3\} \end{Bmatrix}; \\ c(n,a) &= \begin{Bmatrix} 0, \text{ if } f(n,a) \in \{0,2\} \\ 1, \text{ if } f(n,a) \in \{1,3\} \end{Bmatrix}. \end{aligned} \tag{10}$$

Then the integer n is located at vector (i, j), where

$$i = \sum_{a=0}^{N-1} 2^{N-a-1} r(n,a) \quad j = \sum_{a=0}^{N-1} 2^{N-a-1} c(n,a). \tag{11}$$

This ordering is more easily understood with an example. Considered the integer 14 in the second array of Fig. 2. The formulas above, with $N = 2$, give

$$\begin{aligned} &f(14,0) = g_1(14) = h_5(2) = 1 \\ &f(14,1) = g_2(14) + h_1(3) = h_3(2) + h_1(3) = 3 + 3 = 2(\mathrm{mod}\,4). \end{aligned}$$

Thus by (11) the row index for 14 is:

$$2^1 \cdot 0 + 2^0 \cdot 1 = 1$$

and the column index is:

$$2^1 \cdot 1 + 2^0 \cdot 0 = 2.$$

Using the way of table *b, c, d* symmetric and asymmetry matrices 4×4 interleaving, assumed that the position of continuous error occurs is the same as (5), the degree of error dispersion after interleaving can be seen.

$$X'_{1(b)} = \begin{bmatrix} x'_0 & x'_8 & x'_2 & x'_{10} \\ x_{12} & x_4 & x'_{14} & x'_6 \\ x_3 & x_{11} & x_1 & x_9 \\ x_{15} & x_7 & x_{13} & x_5 \end{bmatrix}; \tag{12}$$

$$X_{4(b)} = (x'_0, x_1, x'_2, x_3, x_4, x_5, x'_6, x_7, x'_8, x_9, x'_{10}, x_{11}, x_{12}, x_{13}, x'_{14}, x_{15}); \tag{13}$$

$$X'_{1(c)} = \begin{bmatrix} x'_0 & x'_4 & x'_6 & x'_{10} \\ x_8 & x_{12} & x'_{14} & x'_2 \\ x_{15} & x_3 & x_9 & x_{13} \\ x_7 & x_{11} & x_1 & x_5 \end{bmatrix}; \tag{14}$$

$$X_{4(c)} = (x'_0, x_1, x'_2, x_3, x'_4, x_5, x'_6, x_7, x_8, x_9, x'_{10}, x_{11}, x_{12}, x_{13}, x'_{14}, x_{15}); \tag{15}$$

$$X'_{1(d)} = \begin{bmatrix} x'_0 & x'_3 & x'_6 & x'_8 \\ x_{10} & x_{12} & x'_{14} & x'_1 \\ x_4 & x_7 & x_9 & x_{11} \\ x_{13} & x_{15} & x_2 & x_5 \end{bmatrix}; \quad (16)$$

$$X_{4(d)} = (x'_0, x'_1, x_2, x'_3, x_4, x_5, x'_6, x_7, x'_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x'_{14}, x_{15}). \quad (17)$$

The comparation of these interleaving methods are presented in the table.

Performance comparison of Interleaving methods

Continuous error bit	The number of possible errors	Number of errors that cannot be discrete			
		a	*b*	*c*	*d*
2	15	0	0	0	0
3	14	0	0	2	0
4	13	0	1	3	0
5	12	12	1	4	0
6	11	11	3	6	0
7	10	10	4	6	0
8	9	8	7	6	8

It can be seen from the above table that the *d* method has an ideal performance when faced the continuous errors which is less than 7. However, there is not exist one method which have enough ability to deal with the continued errors which are eight and above.

Conclusion

In this paper, the methods of transforming one-dimensional linear code into two-dimensional symmetric and asymmetric recursive scattering information code were analyzed. The method has transformed the encoded information sequence into a discrete sequence to combat the burst noisy environment. The experiment results indicate that the new method has a good performance.

References

1. Cao X.H., Zhang Z.H. // Information Theory and Coding (Third Edition). 2014–1–134. P. 142–153.
2. Zhang H. Digital Communication Technology [M]. Beijing: Peoples Post and Telecommunications Publishing House, 2008.
3. Shu L., Daniel J.C., Jr. // Error Control Coding, Second Edition (ISBN 0–13–042672–5). 2004. P. 44–89.
4. Bruckstein A.M., Holt R.J., Netravali A.N. // Holographic Representations of Images. IEEE Transaction on Image Processing. 1998. Vol. 7, No. 11. P. 1583–1597.
5. Haralick R.M., Shanmugam K.I. Dinstein. // Textural Features for Image Classification. IEEE Trans. on Systems, Man, and Cybernetics. 1973. Vol. SMC-3, No. 6. P. 610–621.

УДК 004.056.55

УСЕЧЕННЫЙ ПОЛИНОМИАЛЬНО-НОРМЕННЫЙ МЕТОД КОРРЕКЦИИ ОШИБОК В НЕПРИМИТИВНЫХ БЧХ-КОДАХ C_5

Е.В. СЕРЕДА[1], В.А. ЛИПНИЦКИЙ[2], В.К. КОНОПЕЛЬКО[1]

[1]*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь*
[2]*Военная академия Республики Беларусь, Республика Беларусь*

Поступила в редакцию 20 марта 2019

Аннотация. Рассмотрены полиномиальные инварианты *G*-орбит ошибок, которые создают хорошие перспективы для декодирования больших спектров ошибок. Приведена развернутая формулировка усеченного двухуровневого полиномиально-норменного метода коррекции всех допустимых реальным минимальным расстоянием ошибок непримитивными БЧХ-кодами. В цепочку «синдром-ошибка» норменные методы ввели промежуточный уровень идентификации Г-орбиты, которой принадлежит искомая ошибка, что сокращает поисковые процедуры. Данный метод вводит еще один уровень идентификации – определение узкой группы *G*-орбит, среди которых лежит искомый вектор ошибок.

Ключевые слова: БЧХ-код, автоморфизм кодов, группа циклических сдвигов координат векторов, циклотомические подстановки.

Введение

Теория норм синдромов [1–3] обосновала норменный метод декодирования ошибок в семействе БЧХ-кодов, который легко реализуется и действует на порядок быстрее иных синдромных методов. Данный метод может быть усовершенствован и ускорен с помощью полиномиальных инвариантов.

Рассмотрим циклические двоичные БЧХ-коды C длины $n=(2^m-1)/\tau$ для целых $\tau>1$, задаваемые проверочными матрицами вида $H=[\beta^i,\beta^{3i}]^T$, $0\le i\le n-1$, где $\beta^\tau=\alpha$ и α – корень примитивного неприводимого над $GF(2)$ полинома степени m, $m\ge 3$, α – примитивный элемент поля определения кода C, поля Галуа $GF(2^m)$ [2–4]. Из непримитивных БЧХ-кодов C_5 с проверочной матрицей H с практической точки зрения представляют интерес только те коды, минимальное расстояние которых больше конструктивного: $d>5$. Обозначим такие коды C_5^+.

Проблема декодирования в кодах C_5^+

Пусть инфокоммуникационная система (ИКС) на основе БЧХ-кода C_5^+ приняла сообщение $\bar{x}$. Вычисляем синдром ошибок $S(\bar{x})=H\cdot\bar{x}^T=(s_1,s_2)^T$, где s_1, s_2 – элементы поля Галуа $GF(2^m)$. Неравенство $S(\bar{x})\ne\bar{0}$ означает существование ошибок в принятом сообщении: $\bar{x}=\bar{c}+\bar{e}$, где $\bar{c}$ – истинное сообщение, $\bar{e}$ – вектор ошибок. Поскольку $H\cdot\bar{c}^T=\bar{0}$, то $S(\bar{x})=S(\bar{e})$ – зависит только от вектора-ошибки $\bar{e}$.

Группа Г – циклическая группа порядка n – состоит из степеней циклической подстановки σ, действующей на каждый вектор $\bar{x}=(x_1,x_2,...,x_n)$ по правилу: $\sigma(\bar{x})=(x_n,x_1,x_2,...,x_{n-1})$. Для всякой вектор-ошибки $\bar{e}=(e_1,e_2,...,e_n)$ ее Г-орбита $<\bar{e}>_\Gamma=<\bar{e}>=\{\bar{e},\sigma(\bar{e}),...,\sigma^{\nu-1}(\bar{e})\}$, где ν – наименьшее натуральное число с условием: $\sigma^\nu(\bar{e})=\bar{e}$. Известно, что при этом ν делит n или же $\nu=n$. При $\nu=n$ Г-орбита $<\bar{e}>$ называется полной. Действие σ на вектор $\bar{e}$

в БЧХ-коде C с проверочной матрицей H единовременно влечет изменение синдрома: если $S(\bar{e}) = H \cdot \bar{e}^T = (s_1, s_2)^T$, то $S(\sigma(\bar{e})) = H \cdot \sigma(\bar{e})^T = (\beta \cdot s_1, \beta^3 \cdot s_2)^T$ [1–3]. Из последнего равенства следует определение нормы синдрома: $N(S(\bar{e})) = s_2 / s_1^3$. Так как $N(S(\sigma(\bar{e}))) = N(S(\bar{e}))$, то вычисленная норма становится инвариантом – нормой Г-орбиты $J = <\bar{e}>_\Gamma$ – и обозначается $N(J) = N(<\bar{e}>_\Gamma)$.

Отметим также, что могут существовать ошибки $\bar{e}$ веса 3, первая компонента синдрома которых $s_1 = 0$. Тогда норма $N(<\bar{e}>_\Gamma) = \infty$ и не является элементом поля Галуа $GF(2^m)$.

Группа G некоммутативна, имеет порядок mn и получается присоединением к группе Г циклотомической подстановки φ, переставляющей координаты векторов n-мерного (n – нечетно) пространства по правилу: $\varphi(i) = \begin{cases} 2i-1, 2i-1 \le n, \\ 2i-1-n, 2i-1 > n. \end{cases}$ При этом подстановка φ имеет порядок m, $\varphi\sigma = \sigma^2\varphi$. Всякую Г-орбиту $<\bar{e}>$ подстановка φ преобразует в новую Г-орбиту, поэтому для каждого вектора-ошибки $\bar{e}$ БЧХ-кода C ее G-орбита имеет вид: $<\bar{e}>_G = \left\{<\bar{e}>_\Gamma, \varphi(<\bar{e}>_\Gamma), ..., \varphi^{\mu-1}(<\bar{e}>_\Gamma)\right\}$, где μ – наименьшее натуральное число с условием: $\varphi^\mu(<\bar{e}>_\Gamma) = <\bar{e}>_\Gamma$. При этом μ делит m или же $\mu = m$. При $\mu = m$ G-орбита $<\bar{e}>_G$ называется полной при условии, что Г-орбита $<\bar{e}>_\Gamma$ – полная.

Циклотомическая подстановка определена таким образом, что, если синдром $S(\bar{e}) = H \cdot \bar{e}^T = (s_1, s_2)^T$, а норма синдрома $N(S(\bar{e})) = s_2 / s_1^3 = N$, то $S(\varphi(\bar{e})) = (s_1^2, s_2^2)^T$, $N(S(\varphi(\bar{e}))) = N^2$ [2, 3]. Повторное применение подстановки φ к вектору-ошибке приведет к повторному возведению в квадрат ее синдрома и нормы синдрома. Таким образом, вся G-орбита $<\bar{e}>_G$ имеет синхронный образ в спектре норм этих Г-орбит в виде последовательности норм $\{N, N^2, ..., N^{2^{\mu-1}}\}$, где $N^{2^\mu} = N$.

Полиномиально-норменный метод коррекции ошибок в непримитивных БЧХ-кодах C_5^+

Отображение $f: x \to x^2$ в любом поле Галуа $GF(2^m)$ – автоморфизм Фробениуса этого поля. Поэтому множество $N, N^2, ..., N^{2^{\mu-1}}$ при наименьшем натуральном μ с условием $N^{2^\mu} = N$ является полной системой элементов, сопряженных с элементом N. Полином $p_\mu(N, x) = p(x) = (x-N)(x-N^2)\cdot...\cdot(x-N^{2^{\mu-1}})$ является неприводимым над $GF(2)$ и единственным полиномом степени μ, содержащим все корни множества $N, N^2, ..., N^{2^{\mu-1}}$. Назовем полином $p_\mu(N, x)$ полиномиальным инвариантом G-орбиты [5, 6].

Если $N \in GF(2^m)^*$, что соответствует большинству случаев, то $p(N, x)$ несет основную информацию о своей G-орбите, в частности, степень $p(N, x)$ совпадает с количеством принадлежащих ей Г-орбит. Если $N = 0$, а G-орбита состоит из μ Г-орбит, то полагаем $p(N, x) = x^\mu$. Если же $N = \infty$, то полагаем $p(N, x) = 0$.

На этапе формирования поля $GF(2^m)$ для задания конкретного БЧХ-кода в работающей ИКС, с каждым элементом γ поля $GF(2^m)$ необходимо связать минимальный неприводимый над $GF(2)$ полином $p(\gamma, x)$, одним из корней которого является γ. Затем составляется список PG попарно различных полиномиальных инвариантов $p_i(x)$ всех G-орбит ошибок. С каждым $p_i(x)$ должен быть связан соответствующий i-й список p_iG попарно-различных G-орбит G_{ij}, $1 \le j \le \tau_i \le \tau$. По построению каждая G-орбита G_{ij}, $1 \le j \le \tau_i \le \tau$, представляет собой набор $G_{ij}\Gamma$ Г-орбит $<\bar{e}_{ij\chi}>_\Gamma$, $1 \le \chi \le \mu$, с образующими $\bar{e}_{ij\chi}$, $1 \le \chi \le \mu$, синдромы которых

$S(\bar{e}_{ij\chi}) = (s_1^{\chi}, s_2^{\chi})$ таковы, что $N(S(\bar{e}_{ij\chi})) = N_i^{\chi}$ – различные корни полинома $p_i(x)$. Через $p_iG\Gamma$ обозначаем набор всех Г-орбит, формирующих набор всех G-орбит списка p_iG. Через $p_iG\Gamma N^*$ обозначаем срез, часть тех Г-орбит $J \in p_iG\Gamma$, которые имеют норму $N(J) = N^*$ – один из корней полинома $p_i(x)$ [7].

Пусть ИКС приняла сообщение $\bar{x}$. Вычисляем синдром $S(\bar{x}) = (s_1, s_2)^T$, норму синдрома $N = N(S(\bar{x}))$ и определяем полиномиальный инвариант $p(N, x)$. Найденный полином $p(N, x)$ сравниваем с полиномами списка PG. Пусть $p(N, x) = p_{i_0}(x)$ для какого-то конкретного значения $i = i_0$. Из всего списка G-орбит корректируемых ошибок выделяем коротенький список $G_{i_0 j}$, $1 \le j \le \tau_{i_0} \le \tau$, с полиномиальным инвариантом $p_{i_0}(x)$ и список $p_{i_0}G_{i_0 j}\Gamma$ составляющих их Г-орбит $<\bar{e}_{i_0 j\chi}>_{\Gamma}$, $1 \le \chi \le \mu$.

В каждой G-орбите $G_{i_0 j} \in p_iG$, $1 \le j \le \tau_{i_0} \le \tau$, найдется единственная Г-орбита $<\bar{e}_{i_0 j\chi^*}>_{\Gamma}$, $1 \le \chi^* \le \tau_{i_0}$, с образующей $\bar{e}_{i_0 j\chi^*}$, синдром которой $S(\bar{e}_{i_0 j\chi^*}) = (s_1^{\chi^*}, s_2^{\chi^*})$ таков, что равен вычисленному значению N. Искомая вектор-ошибка $\bar{e}$ в сообщении $\bar{x}$ принадлежит одной из Г-орбит списка $p_{i_0}G_{i_0 j}\Gamma N$.

При условии $s_1 \neq 0$ у синдрома $S(\bar{x}) = (s_1, s_2)^T$ следует вычислять разности $\deg(s_1) - \deg(s_1^{\chi^*})$ и делить их на $\tau = (2^m - 1)/n$. Результатом будет единственное целое значение $\chi^* = \chi_0^*$. Если же $s_1 = 0, \chi^* = \chi_0^*, (\deg(s_2) - \deg(s_2^{\chi_0^*}))/3\tau = l\lambda$, находим то единственное λ, для которого целым является число l. В обоих случаях существует единственное целое l из множества $[0, n)$, удовлетворяющее сравнению $\lambda \equiv l (\mathrm{mod}\, n)$. Вектор-ошибка $\bar{e}$ в сообщении $\bar{x}$ находится по формуле: $\sigma^{\lambda}(\bar{e}_{ij^*\chi_0}) = \bar{e}$. Тогда $\bar{c} = \bar{x} + \bar{e}$ – истинное сообщение.

Заключение

Полиномиально-норменный метод сохранил вычисления, присущие норменным методам, а именно вычисление нормы синдрома, а также величины циклического сдвига образующей Г-орбиты до получения искомой вектор-ошибки. Главное отличие данного алгоритма от норменного – в сокращении однообразного поиска нужной нормы в длинном списке норм всех Г-орбит. Следовательно, переборные процедуры сокращаются.

Полиномиально-норменный метод позволяет достаточно равномерно разделить многообразие Г-орбит корректируемой совокупности на небольшие группы G-орбит по значениям их полиномиальных инвариантов. Исходный поиск начинается в полном списке этих полиномов. Дальнейшие поиски и расчеты ведутся внутри достаточно узкой группы Г-орбит с вычисленным полиномиальным инвариантом. Подобная усеченная процедура коррекции ошибок должна найти применение в высокоскоростных системах передачи информации.

TRUNCATED POLYNOMIAL-NORM METHOD OF ERROR CORRECTION IN BCH-CODES C_5

E.V. SEREDA, V.A. LIPNITSKI, V.K. KONOPELKO

Abstract. Polynomial invariants of *G*-orbits of errors, which create good prospects for decoding large error spectra, are considered. A detailed formulation of the truncated two-level polynomial-norm correction method for all errors that are permissible by a minimum distance, by non-primitive BCH-codes is given. The norm methods in the «syndrome-error» chain introduced an intermediate level of identification of the Γ-orbits to which the sought error belongs. This level reduced the search procedures by an order of magnitude. This method introduces another level of identification – the definition of a narrow group of *G*-orbits, where is the sought error vector.

Keywords: BCH-code, code automorphism, group of cyclic shifts of the vector's coordinates, cyclotomic substitution.

Список литературы

1. Конопелько В.К., Липницкий В.А. Теория норм синдромов и перестановочное декодирование помехоустойчивых кодов. Монография.: Едиториал УРСС, 2004.
2. Липницкий В.А., Конопелько В.К. Норменное декодирование помехоустойчивых кодов и алгебраические уравнения.: Издательский центр БГУ, 2007.
3. Липницкий В.А. Теория норм синдромов. Мн.: БГУИР, 2011.
4. Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. Теория кодов, исправляющих ошибки. М.: Связь, 1979.
5. Липницкий В.А., Середа Е.В. // Докл. БГУИР. 2017. № 5 (107). С. 62–69.
6. Липницкий В.А., Середа Е.В. // Материалы XXII Белорусско-Российской научн.-практ. конф. «Комплексная защита информации». Полоцк, 16 – 19 мая 2017 г. С. 117–120.
7. Липницкий В.А., Середа Е.В. // Веснік Гродзенскага дзяржаўнага ўніверсітэта імя Янкі Купалы. Серыя 2. Матэматыка. Фізіка. Інфарматыка, вылічальная тэхніка і кіраванне. 2019 Т. 9. № 1 (2019). С. 118–127.

УДК 621.39

ЗАЩИТА ПАКЕТНОЙ СТРУКТУРЫ СЕТЕЙ ПЕРЕДАЧИ ИНФОРМАЦИИ НА ОСНОВЕ КОДА РИДА-СОЛОМОНА

С.Б. САЛОМАТИН, А.А.С. Аль-ЭЗАЙРДЖАВИ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Предложена схема кодовой защиты от потери или повреждения пакетов с помощью кодирования информации кодом Рида-Соломона (РС) в каналах передачи данных. Представлены алгоритмы исправления ошибок и стираний кодом РС.

Ключевые слова: помехоустойчивый код Рида-Соломона, ошибка стирания, широковещательный канал передачи данных

Введение

Одна из угроз нарушения целостности и отказа в обслуживании связана с потерей пакетов в сетях передачи данных и мультимедийной информации. Известные методы защиты от потери или повреждения пакетов данных используют технику автоматического повтора и кодирования данных избыточными корректирующими кодами [1–4].

В работе рассматривается метод защиты от потери или повреждения пакетов с помощью двумерного кодирования информации пространственно-временным кодом РС в широковещательных многоуровневых каналах передачи данных. В основе метода лежит способность кода РС исправлять независимые ошибки и ошибки стирания.

Коды Рида-Соломона

Код РС над полем Галуа $GF(q)$ можно определить как код, состоящий из всех слов $\mathbf{c}=[c_0,c_1,...,c_{n-1}]$ длины n, для которых выполняются $(d-1)$ уравнений [5, 6]:

$$\sum_{i=0}^{n-1} c_i\gamma_i^r = 0 \quad c_i \in GF(q), r_0 \le r \le r_0 + d - 2, \qquad (1)$$

где r_0, и d – произвольные целые числа (не больше n); $\gamma_0,\gamma_1,...,\gamma_{n-1}$ – различные ненулевые элементы конечного поля, которые еще называют локаторами i-й позиций кодового слова. Последовательность локаторов позиций $\{\gamma_i\}$ в (1) произвольна и может быть задана любым удобным способом.

Циклический код РС длины n, равной любому делителю $(q-1)$ задается выражением (1) и последовательностью локаторов положений вида $\gamma_i = \alpha^i$, где α – элемент порядка n в поле $GF(q)$, $\alpha = \sqrt[n]{1}$.

Циклический сдвиг влево на l позиций последовательности $\omega(x)\gamma_i = \alpha^i$ эквивалентен умножению локаторов на α^l: $\gamma_i\alpha^l = \alpha^{i+l \bmod n}$, где $n \mid (q-1)$.

Если некоторый вектор $\mathbf{c}=[c_0,c_1,...,c_{n-1}]$ удовлетворяет (1), то и после сдвига получаем

$$\sum_{i=0}^{n-1} c_i\gamma_i^r\alpha^{l_r} = (\sum_{i=0}^{n-1} c_i\gamma_i^r)\alpha^{l_r} = 0,$$

так как $\alpha^{l_r} \neq 0$ при любых l и r, $r_0 \le r \le r_0 + d$.

Проверочная матрица циклического кода РС имеет вид:

$$H = \left[\alpha^{(r_0+i)\, j \bmod n} \right],\ i = 0, ..., m-1,\ j = 0, ...,\ n-1.$$

Это позволяет рассматривать умножение **H** на вектор **c** как вычисление значений полинома $c(x) = \sum_i c_i x^i$ в точках $\{\alpha^{r_0+i}\}$, $i = 0, 1, ..., m-1$.

Определение синдрома. Предположим, что на вход декодера поступает последовательность символов $(y_0, y_1, ..., y_{n-1})$, где $y_i = c_i + e_i$, $\{e_i\}$ – элементы вектора ошибок, $e_i \in GF(q)$. Синдромом кода называется вектор $\mathbf{s} = (s_0, s_1, ..., s_{d-2})$, элементы которого определяются из выражения

$$s_j = \sum_i y_i \gamma_i^{r_0+j},\ j = 0, 1, ...,\ d-2. \qquad (2)$$

Полином синдрома определяется как

$$s(x) = \sum_{j=0}^{m-1} s_j x^j = \sum_{j=0}^{m-1} x^j \sum_{i=0}^{n-1} e_i \gamma_i^{r_0+j},\ \ m = n-k.$$

Если предположить, что в принятом векторе имеются t ошибок, расположенных на позициях с номерами $i_1, i_2, \ldots, i_t$, то можно записать уравнение

$$s(x)\sigma(x) + x^m \Theta(x) = \omega(x), \qquad (3)$$

где $\sigma(x) = \prod_{v=1}^{t} \left(x\gamma_{i_v} - 1 \right)$, $\omega(x) = -\sum_{v=1}^{t} e_{i_v} \gamma_{i_v}^{r_0} \sum_{\mu=1, \mu \neq v}^{t} \left(x\gamma_{i_\mu} - 1 \right)$; $\Theta(x) = \sum_{v=1}^{t} e_{i_v} \gamma_{i_v}^{r_0+m} \sum_{\mu=1, \mu \neq v}^{t} \left(x\gamma_{i_\mu} - 1 \right)$.

Полином $\sigma(x)$ называют полиномом локаторов ошибок, так как его корни являются обратными величинами локаторов искаженных позиций. Степень $\sigma(x)$ равна t, если число ошибок $t < m$.

Полином $\omega(x)$ называется полиномом значений ошибок (полином ошибок). Степень $\omega(x)$ всегда меньше степени $\sigma(x)$. Полиномы $\sigma(x)$ и $\omega(x)$ взаимно просты.

Уравнение (3) часто используется в виде модульного сравнения, которое называется ключевым уравнением:

$$s(x) \cdot \sigma(x) = \omega(x) \mod x^m. \qquad (4)$$

Если полиномы $\sigma(x)$ и $\omega(x)$ известны, тогда для определения локаторов ошибок достаточно найти корни полинома $\sigma(x)$ среди обратных величин локаторов позиций кода, т. е. найти все решения уравнения $\sigma\left(x^{-1}\right) = 0$, $\sigma(x) x \in \{\gamma_i\}$. Если $\sigma\left(x^{-1}\right)$ имеет степень t и ровно t различных локаторов кода являются его корнями, то можно считать, что в принятом векторе действительно t ошибок. Если же хотя бы один корень $\sigma\left(x^{-1}\right)$ не является локатором кодового слова, то в принятом слове больше, чем $((d-1)/2)$ ошибок и $\sigma(x)$ не является правильным полиномом локаторов ошибок.

Если локаторы ошибок известны, то значения ошибок равны:

$$e_{i_v} = \frac{-\omega\left(\gamma_{i_v}^{-1}\right)}{\gamma_{i_v}^{r_0} \prod_{\mu=1, \mu \neq v}^{t} \left(\gamma_{i_v}^{-1} \gamma_{i_\mu} - 1 \right)}.$$

Выражение значения ошибки можно преобразовать к виду формулы Форни:

$$e_{i_\nu} = \frac{-\omega\left(\gamma_{i_\nu}^{-1}\right)\gamma_{i_\nu}^{-r_0+1}}{\sigma'\left(\gamma_{i_\nu}^{-1}\right)},$$

где $\sigma'(x) = \sum_{\nu=1}^{t} \gamma_{i_\nu} \prod_{\mu=1,\mu\neq\nu}^{t} \left(x\gamma_{i_\mu} - 1\right)$ – формальная производная.

Алгебраический алгоритм декодирования кода РС. Исходные данные: принятое слово $\mathbf{y} = (y_0, y_1, ..., y_{n-1})$; параметры кода – $n,\ k,\ d,\ m = d-1,\ r_0$, локаторы кода – $\gamma_0, \gamma_1, ...$

1. Вычисление синдрома $s(x)$ принятого слова

$$s(x) = \sum_{j=0}^{m-1} s_j x^j, \quad s_j = \sum_{i=0}^{n-1} y_i \gamma_i^{r_0+j}.$$

2. Решение ключевого уравнения $s(x)\cdot\sigma(x) = \omega(x) \mod x^m$. Если степень $\deg(\sigma(x)) \geq m/2$, то отказ и переход к 5.

3. Определение множества локаторов ошибок $Z : z \in Z$, если $\sigma\left(\gamma z^{-1}\right) = 0$. Если же не все корни являются локаторами позиций, то отказ и переход к 5.

4. Вычисление значений ошибок и исправление искаженных позиций $\hat{y}_z = y_z$, если $z \notin Z$, $\hat{y}_z = y_z - e_z$ для $z \in Z$,

$$e_z = \frac{-\omega\left(\gamma_z^{-1}\right)\gamma_z^{-r_0+1}}{\sigma'\left(\gamma_z^{-1}\right)}.$$

5. Восстановление информационных символов или выдача признака отказа от декодирования.

Режим исправления ошибок и стираний. В режиме стирания фиксируются не сами оценки принятых символов, а их местоположение и им присваивается статус стертого символа. Суть исправления и стирания состоит в том, что после декодирования можно провести восстановление стертых символов, используя алгоритмы интерполяции. Если при декодировании использовать l стертых символов, тогда два кода будут отличаться друг от друга по меньшей мере на $(d-l)$ позиций, где d – кодовое расстояние. Тогда в дополнение к стиранию можно будет исправлять $t_m = \lfloor (d-l-1)/2 \rfloor$ ошибок, где $\lfloor x \rfloor$ – это целая часть числа x. Код может исправлять все комбинации из ν ошибок и l стираний в канале, для которого $2\nu + l < d$.

Для восстановления одного стертого символа необходим только один проверочный символ – для восстановления значения, т.к. позиция стирания принимающей стороне известна. Например, для поля Галуа $GF(256)$ код РС (94, 88) из 8-битовых символов имеет n = 94, k = 88 и может исправить до 3 ошибок (t = 3) и восстановить до 6 стертых символов.

Алгоритм исправления стираний. Предположим, что выполнено f стираний при приеме кодового слова, в котором имеется ν ошибок. Обозначим локаторы ошибок как $X_1 = \alpha^{i_1}$, $X_2 = \alpha^{i_2}$, ..., $X_\nu = \alpha^{i_\nu}$, а стираний – как $Y_{c,1} = \alpha^{j_1}$, $Y_{c,2} = \alpha^{j_2}$, ..., $Y_{c,f} = \alpha^{j_f}$. Декодирование ведется в следующем порядке.

1. Вычисляется полином локаторов стираний $\Gamma(x) = \prod_{l=1}^{f} (1 - Y_{c,l} \cdot x)$.

2. В декодируемом векторе заменяются символы с координатами стираний на нулевые символы. Для нового вектора находится полином синдрома стираний $s(x)$.

3. Определяется модифицированный полином синдрома $SE(x) = (\Gamma(x)[1 + s(x)] - 1) \bmod x^{2t+1}$.

4. Вычисляется полином локаторов ошибок $\sigma(x)$ с использованием алгоритма Берлекемпа-Месси и значения модифицированного полинома $SE_i, i = f+1, ..., 2t$.

5. Определяются корни уравнения $\sigma(x)$=0 и координаты ошибок.

6. Составляется ключевое уравнение $\omega(x)=\sigma(x)[1+SE(x)]\bmod x^{2t+1}$ и определяется полином локаторов ошибок-стираний $\psi(x)=\sigma(x)\Gamma(x)$.

7. Оцениваются значения ошибок и стираний. Значения ошибок вычисляются по формуле:

$$Q_{i_k}=\frac{-X_k\omega(X_k^{-1})}{\psi'(X_k^{-1})},$$

где ψ' – формальная производная.

8. Значения стираний вычисляют по формуле: $F_{i_k}=\frac{-Y_k\omega(Y_k^{-1})}{\psi'(Y_k^{-1})}$.

Пример. Декодируется код РС(7, 3) построенный над полем *GF*(8).

Принимаемый вектор равен

$$y(x)=\alpha^4x^6+\alpha^5x^5+\alpha^2x^4+x^3+\alpha^6x^2+\alpha^5x+\alpha^6=y_6x^6+y_5x^5+y_4x^4+y_3x^3+y_2x^2+y_1x+y_0.$$

Приемник выдал стирания на позициях $j_1=1, j_2=6,$ что позволяет записать полином стираний как $er(x)=er_1\ x+er_2\ x^6$.

Алгоритм декодирования.

Вычисляется $\Gamma(x)=(1-\alpha x)\left(1-\alpha^6x\right)=1+\alpha^5x+x^2.$

Символы на 1 и 6 позициях в $y(x)$ заменяются нулями: $y_c(x)=\alpha^5x^5+\alpha^2x^4+x^3+\alpha^6x^2+\alpha^6,$ после чего вычисляется синдром:

$$S_1=y_c(\alpha)=\alpha,\ S_2=y_c\left(\alpha^2\right)=\alpha,\ S_3=y_c\left(\alpha^3\right)=\alpha,\ S_4=y_c\left(\alpha^4\right)=\alpha^3,$$
$$s(x)=\alpha x+\alpha x^2+\alpha x^3+\alpha^3x^4.$$

Модифицируется полином синдрома:

$$SE(x)=\left[\left(1+\alpha^5x+x^2\right)\left(\alpha x+\alpha x^2+\alpha x^3+\alpha^3x^4\right)-1\right]\ \bmod\ x^5=\alpha^6x+\alpha^4x^2+\alpha^6x^3+\alpha^2x^4.$$

С помощью алгоритма Берлекемпа-Месси получается выражение для полинома локаторов ошибок: $\sigma(x)=1+\alpha^3x.$ Локатор ошибки $X_1=\alpha^3.$ Ошибка расположена на третьей позиции.

Ключевое уравнение имеет вид:

$$\omega(x)=\sigma(x)\left[1+SE(x)\right]\bmod x^5=\left(1+\alpha^3x\right)\left(1+\alpha^6x+\alpha^4x^2+\alpha^6x^3+\alpha^2x^4\right)\bmod x^5=1+\alpha^4x+\alpha x^2+\alpha^2x^3.$$

Тогда $\psi(x)=\left(1+\alpha^2x+\alpha^3x^2+\alpha^3x^3\right)$ и формальная производная $\psi'(x)=\left(\alpha^2+\alpha^3x^2\right).$

Вычисляются значения ошибок и стираний:

$$Q_3=\frac{\alpha^3\omega(\alpha^4)}{\psi'(\alpha^4)}=\alpha^6,\ F_1=\frac{\alpha\omega(\alpha^6)}{\psi'(\alpha^6)}=\alpha^5,\ F_6=\frac{\alpha^6\omega(\alpha)}{\psi'(\alpha)}=\alpha^4.$$

Моделирование алгоритма в программной среде MathLab показало эффективность алгоритма в режиме совместного исправления независимых ошибок и стираний

Схема сетевого кодирования

Построим двумерный блок информационных данных размером $K\cdot k,$ где $k=k'm,$ каждые *m* бит представляются как символы конечного поля $GF\left(2^m\right)$. Кодирование осуществляется в два этапа: внешнее кодирование (по столбцам) и внутреннее кодирование (по строкам).

Внешнее кодирование осуществляется кодом РС над полем $GF\left(2^m\right)$. Каждому столбцу ставится в соответствие кодовое слово $C_1(n_1, K),$ где $n_1=K+r_0,$ r_0 – число избыточных

символов. Блок содержит k' кодовых слов. Каждая строка блока дополняется заголовком из s бит, после чего кодируется внутренним кодом $C_2(n_2, k_2)$. Каждой строке блока ставится в соответствие код, имеющий, r_2 избыточных бит и длину $n = n_2 = s + k + r_2$, где $k = k_2 - s$. Схема формирует K информационных и r_0 избыточных пакетов в блоке.

Система многоуровневого кодирования. Выберем из каждого пакета информации K' символов $(m \times K')$ бит. Рассматривая эти символы как информационные, можно выполнить кодирование внешним кодом. Аналогичным образом следует поступить с K' избыточными символами каждого внешнего кодового слова.

Предположим, что система имеет L уровней, и на каждом i-ом уровне имеются K_i информационных и R_i избыточных пакетов. В схеме защиты выполняются следующие соотношения: $0 \le K_1 \le K_2 \le ... \le K_L$ и $R_1 \ge R_2 \ge ... \ge R_L \ge 0$.

Если длина блока равна $n_i = K_i + R_i$, то система защиты должна использовать базовый код $\mathrm{PC}(n, k)$, параметры которого должны удовлетворять неравенствам: $n \ge \max\limits_{\forall i}\{n_i\}$ и $k \ge \max\limits_{\forall i}\{K_i\}$, $(n-k) \ge \max\limits_{\forall i}\{R_i\}$.

Структура базового кода должна быть согласована со структурами кодов $\mathrm{PC}(n_i, K_i)$.

Пример. Пусть в блоке передаются 48 информационных пакетов и 12 избыточных пакетов. Выберем по 4 байта из каждого информационного пакета и закодируем 192 байта кодом РС (240, 192). Выберем 4 символа из числа избыточных в каждом полученном коде и используем их в качестве избыточных символов для исправления пакетных ошибок. Таким образом, может быть сформирован пакет из 256 кодов слов кода РС.

Заключение

Корректирующая способность схемы защиты зависит от минимальных кодовых расстояний кодов. Пусть d минимальное кодовое расстояние, ν число ошибок и μ число стираний в принятом слове. Тогда выбор d определяется неравенством $d \ge 2\nu + \mu$. Система кодирования может осуществлять дополнительную функцию распределения ключей, учитывая возможности использования кода РС в полилинейных схемах распределения.

PROTECTION OF THE PACKET STRUCTURE OF INFORMATION TRANSMISSION NETWORKS BASED ON REED-SOLOMON CODE

S.B. SALOMATIN, A.A.S. Al EZAYRDJAVI

Abstract. A code protection scheme against packet loss or damage using the Reed-Solomon code (RS) in data transmission channels is proposed. Algorithms for error correction and erasure by the RS code are given.

Keywords: error-proof Reed-Solomon code, erasure error, broadcast data channel.

Список литературы

1. Rizzo L. // ACM Computer Communication Review. 1997. Vol. 27. No. 2. P. 24–36.
2. Xu. Y., Zhang. T. // IEEE Trans on Broadcasting. 2002. Vol. 48. No. 3. P. 237–245
3. Luby M.G. [et. al.] // Information Theory, IEEE Transactions on. 2001. Vol. 47(2). P. 569–584.
4. Guang X., Fu F.W., Zhang Z. // International Symposium on Network Coding, NetCod. 2011. P. 1–6.
5. Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. Теория кодов, исправляющих ошибки. М.: Связь, 1979.
6. Влэдуц С.Г., Ногин Д.Ю., Цфасман М.А. Алгеброгеометрические коды. Основные понятия. М.: МЦНМО, 2003.

УДК 621.391.14

КОРРЕКЦИЯ ЗАВИСИМЫХ ОШИБОК НА ОСНОВЕ ДЕКОДИРОВАНИЯ САМООРТОГОНАЛЬНЫХ БЛОКОВЫХ КОДОВ

Е.Г. МАКЕЙЧИК, А.И. КОРОЛЕВ, В.К. КОНОПЕЛЬКО

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Рассматривается эффективность коррекции зависимых (модульных) ошибок составными самоортогональными блоковыми кодами со скоростью передачи кодов $R \le 1/3$, построенных на основе равномерных самоортогональных сверточных кодов с пороговым алгоритмом декодирования. Установлено, что при сохранении достоинства алгоритма порогового декодирования сверточных кодов обеспечивается увеличение в $\alpha(\alpha \ge 2)$ раз кратности корректируемых ошибок.

Ключевые слова: самоортогональный сверточный код, блоковый код, синдром, порог, порождающий полином, пороговое декодирование.

Введение

Важнейшей проблемой современных инфокоммуникационных сетей связи является обеспечение высокой достоверности передачи информации. Среди множества методов борьбы с канальными ошибками при передаче информации является помехоустойчивое (корректирующее) кодирование. В цифровых системах передачи информации (ЦСПИ) широкое применение получили сверточные коды (СК) и, в частности, равномерные самоортогональные сверточные коды (РССК) с пороговым алгоритмом декодирования. Однако известны только методы построения канальных кодеков на основе самоортогональных блоковых кодов со скоростью передачи кода $R = 1/2$. В работе рассматривается принцип построения канального кодека на основе составного самоортогонального блокового кода, сформированного на основе РССК со скоростью передачи кода $R \le 1/3$. Определяются основные параметры канального кодека.

Общий принцип построения составных помехоустойчивых кодов

Общий принцип построения составных помехоустойчивых кодов рассмотрен в работах [1–7]. Сущность принципа построения составного помехоустойчивого кода состоит в организации дополнительного увеличения разноса (перемежения) информационных символов, участвующих в формировании проверочных (контрольных) символов кода. Достоинство составных помехоустойчивых кодов состоит в увеличении в $\alpha(\alpha \ge 2)$ раз корректирующей способности базовых (исходных) помехоустойчивых кодов. Параметр α носит название коэффициента перемежения информационных символов, которые используются в формировании проверочных символов кода. В соответствии с [5, 7], способ перемежения может быть простой или обобщенный. Сущность простого способа перемежения информационных символов при кодировании состоит в том, что все показатели степеней порождающего полинома $P(x)$ базового (исходного) кода увеличиваются в α раз, а в порождающей и проверочной матрицах базового кода вводится $(\alpha - 1)$ нулевых строк и столбцов соответственно. При обобщенном способе перемежения в порождающей $\mathbf{G}(x)$ и проверочной $\mathbf{H}(x)$ матрицах базового кода вводится $(\alpha - 1)$ нулевых строк и столбов,

задаваемых показателями степеней порождающего полинома $P(x)$. Характеристики составного РССК с параметрами базового РССК: $P(x)=x^m+x^{m-1}+\ldots+1$, n, k, d_0, $t_{\text{испр.}} \le \alpha\cdot(d_0-1)/2$, $t_{\text{обн.}} \le d_0-1$ (n, k, d_0 – соответственно длина кодовой, информационной последовательностей и минимальное кодовое расстояние) будут равны: $P_{\text{с}}(x)=x^{\alpha\cdot m}+x^{\alpha(m-1)}+\ldots+1$, $d_{0\text{с}}=\alpha\cdot d$, $\alpha\cdot d_0$, $t_{\text{испр. с}} \le \alpha\cdot(d_0-1)/2$ и $t_{\text{обн. с}} \le \alpha(d_0-1)$.

Принципиальное отличие РССК от самоортогонального сверточного кода (ССК) состоит в том, что при кодировании информационных символов формируются (n_0-1), $n_0 \ge 2$ проверочных подпотоков, в то время, как ССК формирует один поток проверочных символов, т. е. $(n_0-k_0)=1$.

Для построения блокового кода на основе составного РССК с пороговым алгоритмом декодирования, корректирующего зависимые ошибки кратностью $t_{\text{испр. сс}} \le \alpha\cdot(d_0-v)/2$ бит, необходимо перейти от непрерывного способа кодирования информационных символов РССК к блоковому способу кодирования, который выполняется по следующей методике [4, 6].

1. Для заданной скорости передачи кода $R=1/n_0$ или допустимой избыточности кода $r=(1-R)$ выбираются порождающие полиномы.

2. Показатели степеней порождающих полиномов умножаются на выбранный коэффициент перемежения α.

3. Определяются параметры кода: $k_{\text{с}} \ge 2\cdot m+1, [\text{бит}]$ – минимально допустимое количество передаваемых (кодируемых) информационных символов, где m – максимальная степень порождающих полиномов; $n_{\text{с}}=2\cdot k_{\text{с}}$, [бит] – длина кодовой последовательности; $d_{0\text{с}}=\alpha\cdot d_0=\alpha(J+1)$ – минимальное кодовое расстояние, где J – число ортогональных проверок кода; $t_{\text{испр.с}} \le \alpha(d_0-1)/2, [\text{бит}]$; $t_{\text{обн.с}} \le \alpha(d_0-1), [\text{бит}]$ – кратность корректируемых и обнаруживаемых ошибок соответственно.

4. Скорость передачи блокового кода, сформированного на основе РССК, зависит от параметров базового РССК и способа формирования кодовой последовательности.

Рассмотрим принцип построения и параметры блокового канального кодека на основе составного РССК с пороговым алгоритмом декодирования.

Блоковый канальный кодек на основе составного РССК

Принцип построения блокового канального кодека рассмотрим на примере использования базового РССК со следующими параметрами:

$R=1/n_0=1/5$, $J=12$, $d_0=J+1=12+1=13$, $q_1(x)=1+D^1+D^7$, $q_2(x)=1+D^2+D^6$,

$q_3(x)=1+D^3+D^{11}$, $q_4(x)=1+D^4+D^{13}$, $r=(1-R)\cdot 100\,\%=(1-0{,}2)\cdot 100\,\%=80\,\%$,

$t_{\text{испр}} \le J/2=12/2=6$ бит, $t_{\text{обн}} \le d_0-1=J=12$ бит.

Выбираем коэффициент перемежения информационных символов $\alpha=2$. Далее рассчитываем параметры блокового канального кодека на основе составного РССК

$q_1(x)=1+x^{\alpha\cdot 1}+x^{\alpha\cdot 7}=1+x^2+x^{14}$, $q_2(x)=1+x^{\alpha\cdot 2}+x^{\alpha\cdot 6}=1+x^4+x^{12}$,

$q_3(x)=1+x^{\alpha\cdot 3}+x^{\alpha\cdot 11}=1+x^6+x^{22}$, $q_4(x)=1+x^{\alpha\cdot 4}+x^{\alpha\cdot 13}=1+x^8+x^{16}$,

$J_{\text{с}}=\alpha\cdot J=2\cdot 12=24$,

$d_0=\alpha\cdot J=2\cdot 12=24$.

$k_{1\text{с}}=2\cdot m_{1\text{с}}+1=2\cdot 14+1=29$ бит – длина информационного блока, где m_i – максимальная степень порождающего полинома, $n_{1\text{с}}=2\cdot k_{1\text{с}}=2\cdot 29=58$ бит – длина кодовой комбинации, $k_{2\text{с}}=2\cdot m_{2\text{с}}+1=2\cdot 12+1=25$ бит, $n_{2\text{с}}=2\cdot k_{2\text{с}}=2\cdot 25=50$ бит, $k_{3\text{с}}=2\cdot m_{3\text{с}}+1=2\cdot 22+1=45$ бит,

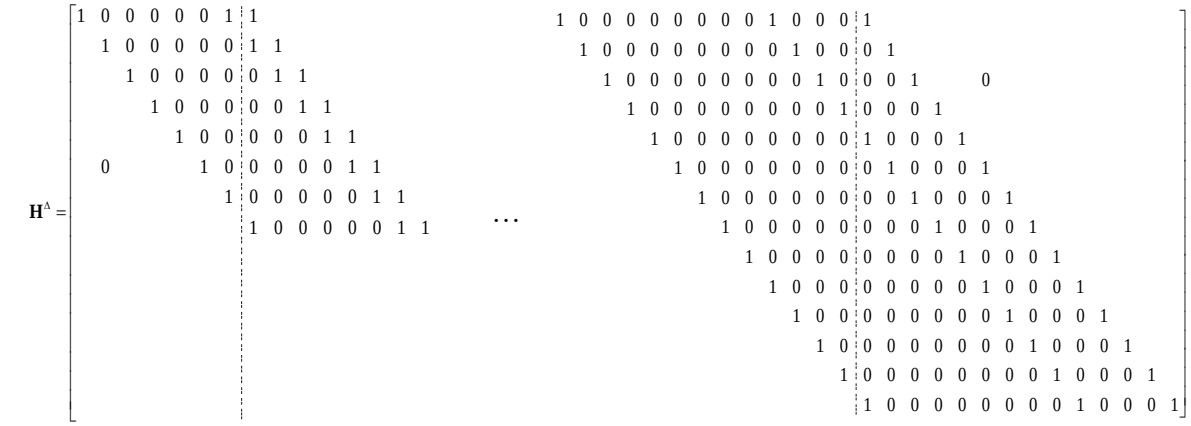
$n_{3c} = 2 \cdot k_{3c} = 2 \cdot 45 = 90$ бит, $k_{4c} = 2 \cdot m_{4c} + 1 = 2 \cdot 26 + 1 = 53$ бита, $n_{4c} = 2 \cdot k_{4c} = 2 \cdot 53 = 106$ бит,

$R_c = k_0 / n_0 = i / i + \Pi_1 + \Pi_2 = 1/3$, $r_c = (1 - R_c) \cdot 100\% = (1 - 0{,}33) \cdot 100\ \% = 67\ \%$,

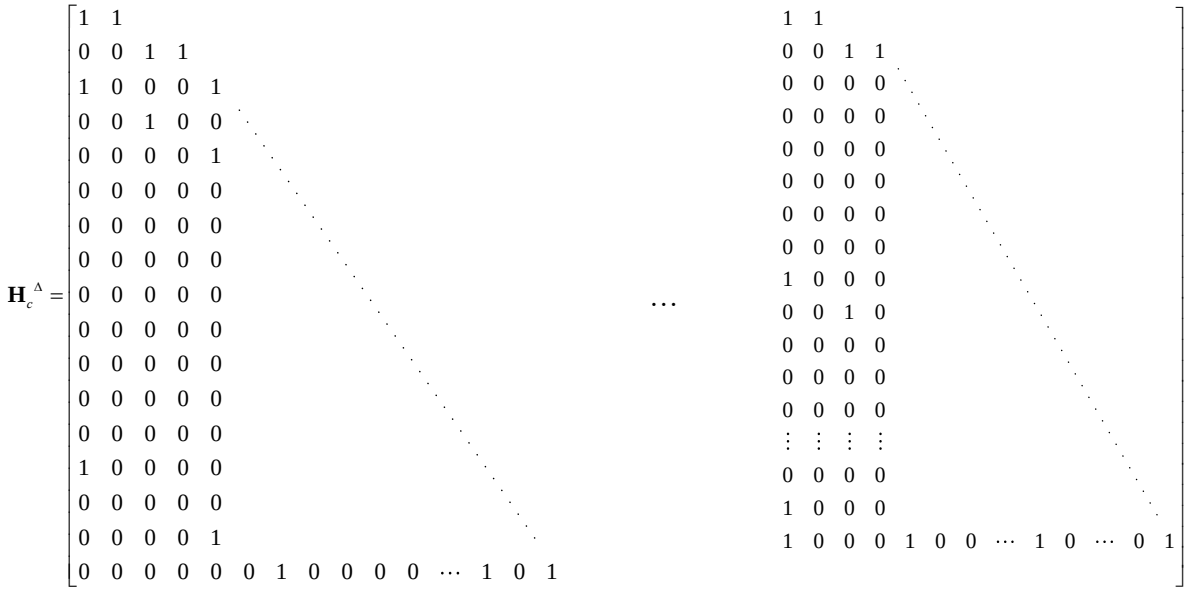
$d_{\text{обн.с}} = \alpha(\alpha_0) = \alpha(J + 1) = 2 \cdot (4 + 1) = 10$, $t_{\text{исп.с}} \le (d_{\text{обн.с}} - 2)/2 = (26 - 2)/2 = 12$ бит,

$t_{\text{обн.с}} \le d_{\text{обн.с}} - 1 = 26 - 1 = 25$ бит.

На рис. 1 приведены проверочные подматрицы базового и составного РССК при реализации блокового способа формирования символов кодовой последовательности РССК.



а



б

Рис. 1. Проверочные подматрицы базового (*а*) и составного (*б*) РССК

На рис. 2 приведены обобщенные структурные схемы канального кодера (*а*) и канального декодера (*б*), выполненных на основе рассмотренных параметров составного РССК, реализующих блоковый способ кодирования и пороговый алгоритм декодирования информационных символов: $\oplus$ – сумматор по модулю два; БУ – буферное устройство, DMX – демультиплексор; МХ – мультиплексор; $\text{ФПС}_\text{к}$ – формирователь проверочных символов кодера; $\text{ФПС}_\text{д}$ – формирователь проверочных символов декодера; КО – корректор ошибок; АСП – анализатор синдромной последовательности; $S(x)$ – синдромная последовательность.

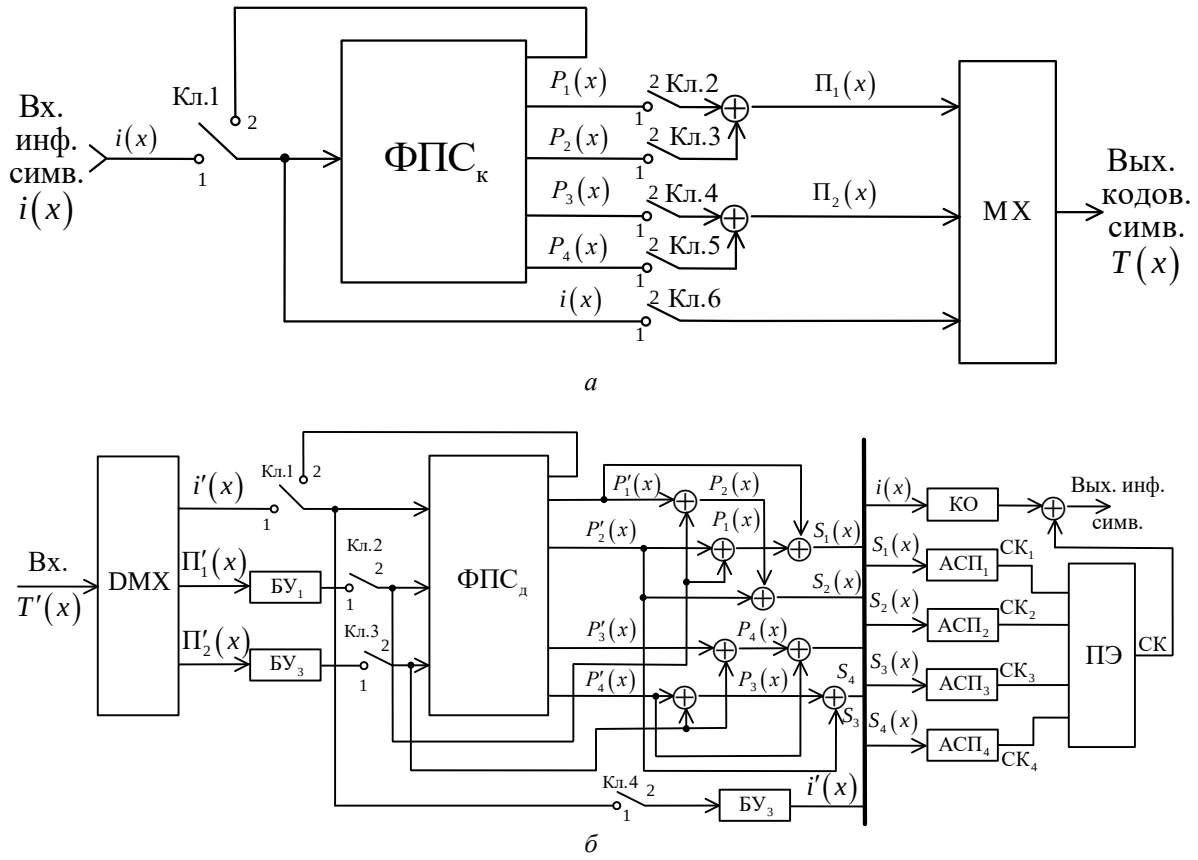


Рис. 2. Обобщенная структурная схема канального кодека на основе составного РССК:
а – кодер; *б* – декодер

Канальный кодер составного РССК содержит регистр сдвига, состоящий из 53 ячеек памяти, и четырех формирователей проверочных символов, выполненных в виде совокупности сумматоров по модулю два, шести ключей и мультиплексора.

Кодирование информационных символов $i(x)$ осуществляется следующим образом. Передаваемые информационные символы записываются в регистр сдвига: ключ Кл. 1 – открыт (положение 1), ключи Кл. 2 – Кл. 6 – закрыты (положение 2). После окончания $k_c = 53$ такта ключ Кл. 1 – закрывается, а ключи Кл. 2 – Кл. 6 – открываются. Далее начинается процесс формирования проверочных символов $P_1 - P_4$ и формирование символов кодовой последовательности $T(x)$. С целью уменьшения избыточности составного РССК осуществляется формирование двух псевдослучайных последовательностей (ПСП) проверочных символов, а именно, $П_1 = P_1(x) \oplus P_2(x)$ и $П_2 = P_3(x) \oplus P_4(x)$. Проверочные символы сформированных ПСП поступают на соответствующие входы МХ, который формирует символы кодовой последовательности путем последовательной передачи информационных символов $i(x)$ и символов $ПСП_1$ и $ПСП_2$.

Из способа формирования символов кодовой последовательности $T(x)$ следует, что скорость передачи составного РССК $R_c = i(x)/i(x) + П_1 + П_2 = 1/3$. Относительная избыточность передаваемой информации $r_c = (1 - R_c) \cdot 100\,\% = (1 - 0{,}33) \cdot 100\,\% = 67\,\%$. Из приведенного расчета следует, что рассматриваемый способ кодирования информации символов эффективнее известного (базового) по избыточности передаваемой информации в $j = r/r_c = 80/67 = 1{,}2$ раза.

Декодирование кодовых символов осуществляется следующим образом. Принятые символы кодовой последовательности $T'(x)$ в DMX распределяются на три подпотока: $i'(x)$,

Π_1 и Π_2. Информационные символы $i'(x)$ через замкнутый ключ Кл. 1 записываются в регистр сдвига $\text{ФПС}_\text{д}$, а символы последовательности Π'_1 и Π'_2 записываются соответственно в буферные устройства БУ_1 и БУ_2. После окончания записи информационных символов ключ Кл. 1 переводится в положение 2, а ключи Кл. 2 – Кл. 4 – в положение 1. Далее осуществляется формирование проверочных символов $P'_1(x) - P'_4(x)$ по правилам $\Pi'_1(x) \oplus P'_1(x)$, $\Pi'_1(x) \oplus P'_2(x)$, $\Pi'_2(x) \oplus P'_3(x)$, $\Pi'_2(x) \oplus P'_4(x)$ и формирование символов синдромных последовательностей $S_1(x) = P_1(x) \oplus P'_1(x)$, $S_2(x) = P_2(x) \oplus P'_2(x)$, $S_3(x) = P_3(x) \oplus P'_3(x)$ и $S_4(x) = P_4(x) \oplus P'_4(x)$.

Сформированные символы синдромных последовательностей поступают на входы соответствующих анализаторов синдромных последовательностей: $\text{АСП}_1 - \text{АСП}_4$. В каждом АСП принимается решение о достоверности полученных информационных символов по принципу (алгоритму) порогового декодирования. Выходной сигнал коррекции (СК) формируется пороговым элементом на основе принятых решений $\text{АСП}_1 - \text{АСП}_4$. Скорректированные информационные символы поступают на выход канального декодера составного РССК.

В соответствии с базовым РССК равной корректирующей способности $(t_{\text{исп}} \le 12 \text{ бит})$ предложенный способ кодирования (декодирования) составного РССК обеспечивает уменьшение длины кодового ограничения в $n_{ac}/n_a = 1054/551 = 2$ раза, где n_{ac} и n_a кодовые ограничения рассчитанные по формуле [5, 7]

$$n_a \ge n_0\left[(n_0 - 1)\cdot J(J-1)/2 + 1\right],$$

где n_0 – длина миниблока кодовых символов, J – число ортогональных проверок.

По энергетическому выигрышу кодирования (ЭВК) предложенный способ составного РССК и способ кодирования/декодирования эффективнее базового РССК в $\beta = \text{ЭВС}_\text{с}/\text{ЭВК} = 10\lg(R_\text{с}\cdot d_{0\text{с}})/10\lg(R\cdot d_0) = 10\lg(3\cdot 26)/10\lg(5\cdot 13) = 1{,}25$ раза.

Заключение

Предложен способ построения блокового канального кодека на основе составного РССК пороговым алгоритмом декодирования со скоростью передачи кода $R \le 1/3$. Установлено, что способ построения канальных кодеков на основе составных РССК, реализующих блоковый способ кодирования и декодирования информационных символов, обеспечивает увеличение корректирующей способности базового (исходного) РССК в $\alpha(\alpha \ge 2)$ раз на меньшей длине кодового ограничения по сравнению с базовым РССК равной корректирующей способности.

CORRECTION OF DEPENDENT ERRORS BASED ON DECODING OF SELF-ORTHOGONAL BLOCK CODES

E.G. MAKEICHIK, A.I. KOROLEV, V.K. KONOPELKO

Abstract. The efficiency correction of dependent (modular) errors by composite self-orthogonal block codes with a transfer rate of codes constructed on the basis of uniform self-orthogonal convolutional codes (RCCS) with a threshold decoding algorithm is considered. It is established that while preserving the merits of the threshold decoding algorithm of convolutional codes, the magnification of the corrected errors increases.

Keywords: self-orthogonal convolutional code, block code, syndrome, threshold, generating polynomial, threshold decoding, code sequence.

Список литературы

1. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. М., 1976.
2. Радченко А.Н., Мирончиков Е.Г. // Радиотехника и электроника. 1961. № 11. С. 18–33.
3. Блох Э.Л., Зяблов В.В. Линейные каскадные коды. М., 1982.
4. Дмитриев О.Ф. // Радиотехника. 1964. Т. 19. № 4. С. 68–75.
5. Касагин Т., Токура Н., Ивадари Е. Теория кодирования. М.,1978.
6. Meggit I.E. // IRE Transaction on Information Theory. 1961. № 4. P. 234–244.
7. Кларк Дж. мл., Кейн Дж. Кодирование с использованием ошибок в системах цифровой связи. Минск, 1987.

УДК 621.39

КОМБИНИРОВАННОЕ РАЗНОСТНОЕ, АРИФМЕТИЧЕСКОЕ И RLE-КОДИРОВАНИЕ БИТОВЫХ ПЛОСКОСТЕЙ ИЗОБРАЖЕНИЙ

Б.Д.С. САДИК, О.М.Х. АЛЬМИЯХИ, М.Н. БОБОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Разработана структура и исследована эффективность комбинированного кодека сжатия изображений без потерь в пространственной области, использующего алгоритмы разностного, арифметического и RLE-кодирования для компактного представления битовых плоскостей изображений.

Ключевые слова: комбинированное кодирование, сжатие изображений без потерь, разностное кодирование, арифметическое кодирование, кодирование длин серий.

Введение

В современных кодеках сжатия изображений [1–6] используется, как правило, один алгоритм обработки для всех битовых плоскостей. Исключением являются кодеки EZW, SPIHT, SPECK [7–9], в которых используется раздельное кодирование битовых плоскостей и предусматривается возможность использования младших битовых плоскостей с равновероятным повтором нулей и единиц без кодирования, и кодек JPEG 2000 [6], в котором предусматривается возможность использования двух алгоритмов для сжатия изображений без потерь и с потерями. Однако в рассмотренных кодеках не предусмотрено применение разных алгоритмов кодирования к различным битовым плоскостям. Эффективность раздельного кодирования битовых плоскостей анализируется в работах [10, 11], но только для кодера RLE [12]. Результаты исследования эффективности использования комбинации арифметического и RLE-кодирования приведены в работе [13], но только для нескольких типов изображений. Часто разностное кодирование используется в комбинации с арифметическим кодированием. Недостаточное исследование комбинированного подхода к кодированию изображений связано с ростом вычислительной сложности из-за дополнительных затрат на выбор алгоритма кодирования и на память для реализации нескольких алгоритмов кодирования. С развитием элементной базы рост вычислительной сложности кодирования становится менее критичным в сравнении с увеличением коэффициента сжатия.

Цель работы – исследование эффективности комбинированного кодека сжатия изображений различных типов без потерь в пространственной области на основе алгоритмов разностного, арифметического и RLE-кодирования.

Комбинированное эффективное кодирование битовых плоскостей изображений

Для повышения коэффициента сжатия изображений без потерь предлагается использовать комбинированное эффективное кодирование на основе разностного, арифметического и RLE-алгоритмов. Сущность подхода состоит в формировании разностных кодов по строкам и столбцам изображений и использовании для сжатия старших битовых плоскостей разностей или их комбинаций нескольких кодеров различных типов $\left(f_{\text{RLE}}\left(B(r)\right)\right.$ и $\left.f_{\text{AC}}\left(I_C\left(r_{HL}, r_{HH}\right)\right)\right)$, лучше учитывающих распределение их значений, и непосредственном включении в результат кодирования младших битовых плоскостей разностей $\left(f_{\text{NC}}\left(B(r)\right)\right.$ при $\left.r \geq 0\right.$ или $\left.f_{\text{NC}}\left(I_C\left(0, r_{LH}\right)\right)\right.$ при $\left.r_{LH} \geq 0\right)$, кодирование которых не эффективно $\left(\left\langle f_{\text{RLE}}\left(B(r)\right)\right\rangle \geq YX\right.$ при $\left.r \geq 0\right.$

или $f\left\langle_{\text{AC}}\left(I_C\left(0,r_{LH}\right)\right)\right\rangle \ge \left(r_{LH}+1\right)YX$ при $r_{LH}\ge 0\big)$ (рис. 1), где $B\left(r\right)$ – битовые плоскости; f_{NC} – функция непосредственного переноса младших битовых плоскостей без кодирования.

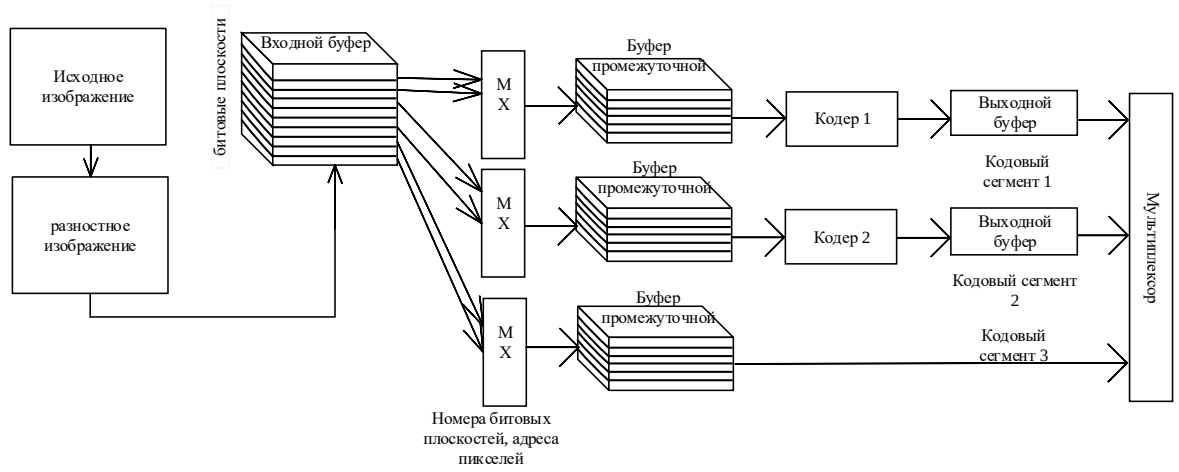


Рис. 1. Схема комбинированного эффективного кодирования битовых плоскостей изображения

Битовые плоскости $B\left(r\right)$ формируются из одинаковых разрядов r разностей $d\left(R+3,y,x\right)$ пикселей $i\left(R,y,x\right)$ R-разрядного изображения $I\left(R\right)=\left\|i\left(R,y,x\right)\right\|_{\left(y=\overline{0,Y-1},\,x=\overline{0,X-1}\right)}$ (3 дополнительных разряда необходимы для учета знаков разностей и увеличения абсолютных значений разностей) и представляют собой матрицу $B\left(r\right)=\left\|b\left(r,y,x\right)\right\|_{\left(y=\overline{0,Y-1},\,x=\overline{0,X-1}\right)}$, состоящую из нулей и единиц $\left(b\left(r,y,x\right)=\left\{0,1\right\}\right)$, размер $Y\times X$ которой совпадает с размером $Y\times X$ исходного изображения $I\left(R\right)$. Старшая битовая плоскость $B\left(R+2\right)$ содержит знаки (0 – обозначает положительные числа, 1 – отрицательные). Значения $d\left(R,y,x\right)$ и $b\left(r,y,x\right)$ связаны выражением:

$$d\left(R+3,y,x\right)=\begin{cases}\sum\limits_{r=0}^{R+2}2^{r}b\left(r,y,x\right) \text{ при } b\left(R+2,y,x\right)=0,\\ -\sum\limits_{r=0}^{R+2}2^{r}b\left(r,y,x\right) \text{ при } b\left(R+2,y,x\right)=1,\end{cases}$$

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

Комбинация из нескольких r_C битовых плоскостей разностей $\left(0<r_C<R+3\right)$ с r_L по r_H $\left(r_H>r_L,\ r_C=r_H-r_L+1\right)$ представляет матрицу $I_C\left(r_L,r_H\right)=\left\|i_C\left(r_L,r_H,y,x\right)\right\|_{\left(y=\overline{0,Y-1},\,x=\overline{0,X-1}\right)}$, значения $i_C\left(r_L,r_H,y,x\right)$ которой имеют r_C разрядов. В этом случае

$$i_C\left(r_L,r_H,y,x\right)=\sum_{r=r_L}^{r_H}2^{r-r_L}b\left(r,y,x\right) \quad \text{при } y=\overline{0,Y-1},\ x=\overline{0,X-1}.$$

Для частного случая $r_H=r_L$ справедливо равенство $i_C\left(r_L,r_H,y,x\right)=b\left(r_L,y,x\right)$.

В табл. 1, 2 приведены средние коэффициенты сжатия битовых плоскостей разностей 8-разрядных $\left(R=8\right)$ спутниковых, портретных, медицинских полутоновых и ландшафтных тепловизионных изображений (примеры приведены на рис. 2), полученные с использованием RLE и арифметического кодеров (AC). Усреднение значений коэффициентов сжатия произведено по 8-ми тестовым изображениям каждого типа.

Таблица 1. **Коэффициенты сжатия битовых плоскостей разностных кодов изображений для кодера RLE**

Битовые плоскости	Частичные и полные коэффициенты сжатия битовых плоскостей разностей изображений		
	Полутоновые спутниковые	Полутоновые портретные	Полутоновые медицинские
8	0,23	0,28	0,35
7	3518,71	2048,00	6553,60
6	116,72	16,86	178,09
5	9,93	2,78	8,52
4	2,36	0,94	1,89
3	0,58	0,46	0,78
2	0,36	0,24	0,57
1	0,40	0,20	0,27
0	0,40	0,20	0,25

Таблица 2. **Коэффициенты сжатия битовых плоскостей разностных кодов изображений для арифметического кодера**

Битовые плоскости	Частичные (полные) коэффициенты сжатия битовых плоскостей разностей изображений		
	Полутоновые спутниковые	Полутоновые портретные	Полутоновые медицинские
8	0,97	0,97	1,00
7	29,75	24,60	27,15
6	27,09	12,49	24,22
5	13,28	5,03	8,92
4	4,51	2,48	3,43
3	1,60	1,43	2,27
2	1,02	1,04	1,49
1	0,97	0,97	1,03
0	0,97	0,96	0,96

Кодер RLE применен отдельно для каждой битовой плоскости разностей $\left(f_{\text{RLE}}\left(B(r)\right)\right.$ при $\left.r=\overline{0,R+2}\right)$, где f_{RLE} – функция RLE-кодирования. На основе получаемого при этом объема $\left\langle f_{\text{RLE}}\left(B(r)\right)\right\rangle$ кода (в битах) вычисляются частичные коэффициенты сжатия с помощью выражения:

$$CR_{\text{RLE}}(r)=YX\Big/\left\langle f_{\text{RLE}}\left(B(r)\right)\right\rangle, \text{ при } r=\overline{0,R=2},$$

где $\langle\ \rangle$ – оператор вычисления объема кода.

а *б* *в*

Рис. 2. Примеры тестовых изображений

Кодирование тестовых изображений целиком (без разделения на битовые плоскости) с помощью кодера RLE не приводит к сжатию $\left(\left\langle f_{\text{RLE}}\left(I(R)\right)\right\rangle \geq RYX\right)$.

Арифметический кодер применен отдельно для старших $\left(f_{\text{AC}}\left(i_C\left(r_{HL}, r_{HH}, y, x\right)\right)\right)$ и младших $\left(f_{\text{AC}}\left(i_C\left(r_{LL}, r_{LH}, y, x\right)\right)\right)$ битовых плоскостей $\left(r_{HH} > r_{HL},\, r_{LH} = r_{HL} - 1,\, r_{LH} > r_{LL},\, r_{LL} \geq 0\right)$, где f_{AC} – функция арифметического кодирования. Частичные коэффициенты сжатия вычисляются с помощью выражений

$$CR_{\text{AC}}\left(r_{HL}, r_{HH}\right) = \left(r_{HH} - r_{HL} + 1\right)YX \Big/ \left\langle f_{\text{AC}}\left(I_C\left(r_{HL}, r_{HH}\right)\right)\right\rangle,$$

$$CR_{\text{AC}}\left(r_{LL}, r_{LH}\right) = \left(r_{LH} - r_{LL} + 1\right)YX \Big/ \left\langle f_{\text{AC}}\left(I_C\left(r_{LL}, r_{LH}\right)\right)\right\rangle.$$

В табл. 3 приведены наиболее эффективные комбинации арифметического и RLE кодеров для битовых плоскостей разностных кодов тестовых изображений, и коэффициенты сжатия CR_{C}, соответствующие этим комбинациям.

Таблица 3. **Наиболее эффективные комбинации арифметического и RLE кодеров для разностных кодов изображений**

Битовые плоскости	Комбинации арифметического и RLE кодеров для разностных кодов изображений		
	Полутоновые спутниковые	Полутоновые портретные	Полутоновые медицинские
8	NC	NC	NC
7	RLE	RLE	RLE
6	RLE	RLE	RLE
5	AC	AC	AC
4	AC	AC	AC
3	AC	AC	AC
2	AC	AC	AC
1	NC	NC	AC
0	NC	NC	NC
CR_{C}	2,05	1,85	2,29
CR_{AC}	2,01	1,82	2,23
CR_{RLE}	1,76	1,48	1,72

Коэффициент сжатия CR_{C} при комбинированном эффективном кодировании вычисляется с помощью выражения

$$CR_{\text{C}} = \frac{RYX}{\sum_{m=1}^{M}\left\langle f_{\text{RLE}}\left(B\left(r\left(m\right)\right)\right)\right\rangle + \sum_{n=1}^{N}\left\langle f_{\text{AC}}\left(I_C\left(r_L\left(n\right), r_H\left(n\right)\right)\right)\right\rangle + \sum_{p=0}^{P-1}\left\langle f_{\text{NC}}\left(B\left(p\right)\right)\right\rangle}$$

при $R = M + N + P,\quad r_H\left(n\right) > r_L\left(n\right),\quad M \geq 0, N \geq 0, P \geq 0,$ где M – число битовых плоскостей, кодируемых с помощью кодера RLE; N – число битовых плоскостей, кодируемых с помощью арифметического кодера; P – число битовых плоскостей, непосредственно переносимых в результирующий код.

Исходя из табл. 3, для повышения коэффициента сжатия предлагается следующее правило комбинированного кодирования разностей пикселей изображений, определяющее формирование результирующего кода C_{C}: необходимо использовать арифметический кодер для старшей знаковой плоскости и битовых плоскостей 5–2, кодер RLE – для битовых плоскостей 6–7, перенос без кодирования – для битовых плоскостей 8 и 1–0.

Из табл. 3 следует, что комбинированное разностное, арифметическое и RLE-кодирование позволяет повысить коэффициент сжатия изображений в среднем в $CR_{\text{C}}/CR_{\text{AC}}$ =1,02 и $CR_{\text{C}}/CR_{\text{RLE}}$ =1,25 раза по сравнению с арифметическим и RLE-кодированием соответственно.

Заключение

Установлена эффективность комбинированного разностного, арифметического и RLE-кодирования для сжатия изображений. Выбор оптимальных комбинаций битовых плоскостей разностных кодов изображений для арифметического и RLE кодирования позволяет повысить коэффициент сжатия в среднем в 1,02 и 1,25 соответственно.

COMBINED DIFFERENCE, ARITHMETIC AND RLE-CODING OF IMAGE BIT PLAYS

B.S.S. SADIK, O.M.Kh. ALMIYAHI, M.N. BOBOV

Abstract. A structure and efficiency of combined lossless image compression codec in the spatial domain using arithmetic and RLE-coding algorithms for compact representation of bit planes of the image was developed and researched.

Keywords: combined coding, lossless image compression, differential coding, arithmetic coding, run length coding.

Список литературы

1. Ватолин Д. [и др.] Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео М: Диалог-МИФИ, 2003.
2. Сэломон Д. Сжатие данных, изображений и звука М: Техносфера, 2004.
3. Тропченко А.Ю., Тропченко А.А. Методы сжатия изображений, аудиосигналов и видео СПб: СПбГУ ИТМО, 2009.
4. Гонсалес Р. Вудс Р. Цифровая обработка изображений М: Техносфера, 2006.
5. Ziv J., Lempel A. // IEEE Transactions on Information Theory, IT-23. 1977. P. 337–343.
6. Taubman D.S., Marcellin M.W. JPEG2000: image compression fundamentals, standards, and practice Springer–Verlag, 2002.
7. Shapiro J.M. // IEEE Trans. Signal Processing. 1993. № 41. P. 3445–3462.
8. Said A., Pearlman W.A. // IEEE Trans. on Circuits and Systems for Video Technology. 1996. Vol. 6. P. 243–250.
9. Islam A., Pearlman W.A. // ISO/IEC/JTC1/SC29, WG1. 1998. № 873. P. 312–326.
10. Аль-Бахдили Х.К. [и др.] // Докл. БГУИР. 2016. № 2 (96). С. 63–69.
11. Al-Bahadily H.K. [et al.] // International Journal of Advanced Computer Science and Applications. 2016. Vol. 7, No. 7. P. 250–255.
12. Golomb S.W. // IEEE Transactions on Information Theory. July, 1966. P 399–401.
13. Abdmouleh M.K., Masmoudi A., Bouhlel M.S. // Journal of Software Engineering and Applications. 2012. No. 5. P. 41–44.

УДК 621.391(075.8)

AN ALGORITHM FOR DECODING PRODUCT CODES BASED ON SOFT-DECISION DECODING OF DUAL CODES OF THE COMPONENT CODES

PHAM KHAC HOAN[1], NGUYEN TRUNG THANH[1],
NGUYEN THI HONG NHUNG[2], NGUYEN ANH TUAN[3]

[1]*Le Quy Don Technical University, Ha Noi, Viet Nam*
[2]*University of Economic and Technical Industries, Ha Noi, Viet Nam*
[3]*Belarusian State University of informatics and radioelectronics, Republic of Belarus*

Submitted 20 March 2019

Abstract. Product codes have large minimum Hamming distance and their complexity might be compared with Turbo codes. Conventional algorithms for decoding product codes have low decoding performances and very high complexities. In order to ensure the applicability of product codes in practice, this article proposes a method for decoding product codes which provide good decoding performance and acceptable complexity. Simulation results of the proposed algorithm show that, a signal-to-noise ratio of about only 4,1 dB is required in order to produce a bit error ratio (BER) not to exceed10^{-6} with rate code at 0,7034.

Keywords: product code, concatenated code, block turbo code, iterative decoding, SDDDCA-PC.

Introduction

Recently, coding theory has made great progress and has found good codes that be capable of closely approaching the Shannon capacity limit [1]. The Turbo code based on concatenation of small length component codes is one of mostly used codes. Although structure of the first concatenated code introduced was the block turbo codes (BTC), in practice, researchers mainly focus on turbo codes under the convolutional turbo codes (CTC) [2]. The reason that not much attention is paid to the BTC is the decoding performance is poor. If the same decoding performance to both the CTC and BTC to be achieved, the BTC has greater complexity.

Product codes are serial concatenated codes, and constituted of component codes which are short length block codes, firstly introduced by Elias in 1954 [3]. Product codes have large error-correction capability due to large minimum Hamming distance, however, decoding algorithms are complicated.

The iterative (turbo) decoding can be applied to product codes, using the maximum a posterior probability (MAP) to component codes. If the turbo decoding is applied to product codes, in block codes, the maximum number of vertices at each time in the trellis will be equivalent (use the famous Wolf bound [4]) to the total code words of linear code or dual code of a linear code. This is a major obstacle to the use of good block codes instead of convolutional codes in the turbo decoding concatenated codes. Obviously, the reduction in the complexity of the MAP decoding of block codes really decreases the total complexity of turbo decoding for product codes. One of the possible solutions is to increase the number of iterations in the turbo decoding to compensate for the use of the non-optimum MAP decoder for the component codes. However, as the number of iterations increases, the decoding delay increases as well. Many proposals have been implemented in order to solve the complex problem of MAP decoding for component codes.

In 1996, Hagenauer studied the turbo decoding for product codes with the MAP decoder for component codes [5]. Hagenauer proposed an optimum decoding algorithm, named soft output viterbi algorithm (SOVA), which was similar to the MAP decoding using dual codes. Although this decoding method is optimal, but it is only suitable for codes with high coding rate. Meanwhile, for linear block codes with high coding rate, it is difficult to achieve a sufficient minimum Hamming distance for modern communication applications. Moreover, the use of nonlinear functions results in very high

complexity, making this method is useful for the theoretical research but difficult for practical applications [5, 6].

Later, further researches tried to reduce the complexity of the turbo decoding method for product codes, by trading off decoding time against performance [7, 8]. In 1998, Pyndiah suggested another approximation for the MAP decoding of the component codes. In this algorithm decoding a Chase type II decoder was used to obtain a list of codewords, those are the closest to the received sequence, then the MAP decoder is implemented using only a subset of those codewords instead of the whole codewords [9, 10]. This algorithm produces a decoding performance equivalent to turbo decoding with Hagenauer's MAP decoding for component codes. However, Pyndiah's approximation cannot always be explained or presented with a theoretical basis. This makes his algorithm difficult to analyze, and therefore, it is not feasible to be neither improved nor applied to other codes.

One of the new effective decoding methods is to scan all the decoding message in the dual code set of the original code set. Derived from this research direction, the article proposes a new decoding method for product codes. This method is based on soft-decision decoding of dual codes of the component codes, which are block codes with small redundancy. For codes with small redundancy, the decoding will reduce the complexity while the decoding messageremains the same as the original code. This is because the number of codewords in the dual code of the high-speed codes is much less than that of the original code [11]. This suggestion can avoid the MAP decoding of component codes, which may contribute to exploitthe capacity of error control of product code in new generation communication systems. The remainder of this article is organized as follows: Section II reviews encoding of communication system using product codes and proposes a novel algorithm with sufficient theoretical basis. Section III presents a simulation results of the proposed algorithm on a binary input additive white Gaussian noise (AWGN) channel. Finally, the conclusion is given in section IV.

The soft-decision of dual codes decoding algorithm of the component codes of product codes

1. Encoding product codes. Let $C_1(n_1,d_1,k_1)$ and $C_2(n_2,d_2,k_2)$ are linear block codes with the generator matrices $\mathbf{G}_1$, $\mathbf{G}_2$, and with the parity check matrices $\mathbf{H}_1$, $\mathbf{H}_2$, respectively.

In the encoding process, $k_1 \times k_2$ bits of message are coded to codeword with $n_1 \times n_2$ bits, the code rate is (k_1 / n_1), (k_2 / n_2) and minimum Hamming distance is $(d_1 \times d_2)$, in which d_1 and d_2 are minimum Hamming distance of C_1 and C_2 respectively. Fig. 1 illustrates the structure of product code, $C = C_1 \otimes C_2$.

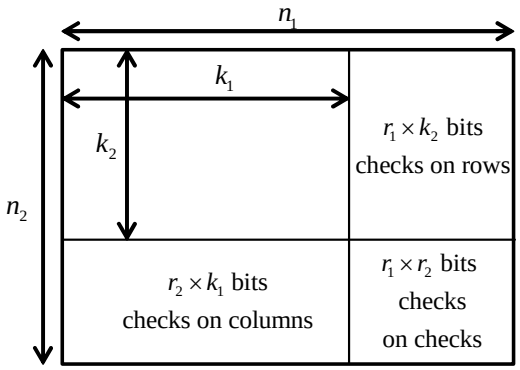


Fig. 1. Construction of product codes

At the input of the coder message u (size $k_1 \times k_2$) is coded by product code C with generator matrix $\mathbf{G}$ to produce codeword c (size $n_1 \times n_2$). A codeword c in the product code can either be generated by multiplying a $k_1 \times k_2$ long binary vector with generator matrix for C or by using the following equation:

$$c = \mathbf{G}_2^T \otimes u \otimes \mathbf{G}_1 \quad (1)$$

in which $\mathbf{G}_2^T$ is the transposed of the matrix $\mathbf{G}_2$; $\otimes$ is the Kronecker product.

Codeword c is then modulated BPSK. Suppose this codeword is modulated to binary signal ± 1 with rule $x = 1 - 2c$ and is transmitted over the Gaussian channel with zero mean and variance $2\sigma^2$. The received signal is:

$$y = x + w, \tag{2}$$

in which $w = (w_1, w_2, w_3, ..., w_n)$ is the noise vector and $y_m = x_m + w_m,\ 1 \le m \le n.$

2. Theoretical basis of the soft-decision decoding algorithm of dual codes. This section poposes the theoretical basis of the soft-decision decoding algorithm for dual codes of the component codes of product codes.

Let C^* is the dual code of C, and $c_j^* = (c_{j1}^*, c_{j2}^*, ..., c_{jn}^*)$ is its j-th codeword. Denote $P(y_m \mid i),\ i \in \{0,1\}$ as the conditional probability of the event obtaining y_m when bit node $c_m = i$ is sent and denote $\varphi_m = P(y_m \mid 1) / P(y_m \mid 0)$ as the likelihood ratio of m-th bit. It is possible to deduce $\varphi_m = \exp(-2y_m / \sigma^2)$. Let following equations:

$$A_m(0) = \sum_{t=0}^{1} \sum_{j=1}^{2^{n-k}} \prod_{l=1}^{n} \sum_{i=0}^{1} (-1)^{i(c_{jl}^* + t\delta_{ml})} P(y_m \mid i), \tag{3}$$

$$A_m(1) = \sum_{t=0}^{1} (-1)^t \sum_{j=1}^{2^{n-k}} \prod_{l=1}^{n} \sum_{i=0}^{1} (-1)^{i(c_{jl}^* + t\delta_{ml})} P(y_m \mid i), \tag{4}$$

In [5] it was proved that $A_m(0) = \lambda P(0 \mid y)$ and $A_m(1) = \lambda P(1 \mid y),$ with λ is a positive constant. It's possible to obtain:

$$\frac{A_m(1)}{A_m(0)} = \frac{P(1 \mid y)}{P(0 \mid y)} = \frac{P(1 \mid y_m) P(y_m \mid y)}{P(0 \mid y_m) P(y_m \mid y)} = \frac{P(y_m \mid 1) P(1)}{P(y_m \mid 0) P(0)} = \frac{P(y_m \mid 1)}{P(y_m \mid 0)} = \varphi_m, \tag{5}$$

with an assumption that bit 0 and 1 are sent with a same probability.

Therefore, according to equation (5), proposed decoder can decide output message based on the value of the likelihood ratio of posterior probability

$$\begin{cases} \varphi_m \ge 1 \to c_m = 1 \\ \varphi_m < 1 \to c_m = 0 \end{cases}$$

and the value of φ_m can be the decoding message for the next decoder.

On the other hand:

$$\frac{A_m(1)}{A_m(0)} = \frac{[A_m(0) + A_m(1)] - [A_m(0) - A_m(1)]}{[A_m(0) + A_m(1)] + [A_m(0) - A_m(1)]}$$

And from [5]:

$$A_m(0) - A_m(1) = 2 \sum_{j=1}^{2^{n-k}} \prod_{l=1}^{n} \sum_{i=0}^{1} (-1)^{i(c_{jl}^* + \delta_{ml})} P(y_m \mid i) = \lambda \sum_{j=1}^{2^{n-k}} \prod_{l=1}^{n} \left(\frac{1 - \varphi_l}{1 + \varphi_l} \right)^{c_{jl}^* \oplus \delta_{ml}}. \tag{6}$$

From equation (3) and (4) it's possible to deduce:

$$A_m(0)+A_m(1)=\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\sum_{i=0}^{1}(-1)^{i(c_{jl}^*)}P(y_m\,|\,i)+\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\sum_{i=0}^{1}(-1)^{i(c_{jl}^*+\delta_{ml})}P(y_m\,|\,i)+$$

$$+\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\sum_{i=0}^{1}(-1)^{i(c_{jl}^*)}P(y_m\,|\,i)-\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\sum_{i=0}^{1}(-1)^{i(c_{jl}^*+\delta_{ml})}P(y_m\,|\,i)= \quad (7)$$

$$=2\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\sum_{i=0}^{1}(-1)^{i(c_{jl}^*)}P(y_m\,|\,i)=\lambda\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\left(\frac{1-\varphi_l}{1+\varphi_l}\right)^{c_{jl}^*}$$

Therefore, the value of φ_m can be calculated as:

$$\frac{A_m(1)}{A_m(0)}=\varphi_m=\frac{\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\left(\frac{1-\varphi_l}{1+\varphi_l}\right)^{c_{jl}^*}-\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\left(\frac{1-\varphi_l}{1+\varphi_l}\right)^{c_{jl}^*\oplus\delta_{ml}}}{\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\left(\frac{1-\varphi_l}{1+\varphi_l}\right)^{c_{jl}^*}+\sum_{j=1}^{2^{n-k}}\prod_{l=1}^{n}\left(\frac{1-\varphi_l}{1+\varphi_l}\right)^{c_{jl}^*\oplus\delta_{ml}}}. \quad (8)$$

3. Decoding product codes based on the soft-decision decoding of dual codesof the component codes. Fig. 2 illustrates the decoding process of the soft-decision decoding of dual codes of the component codes for product codes.

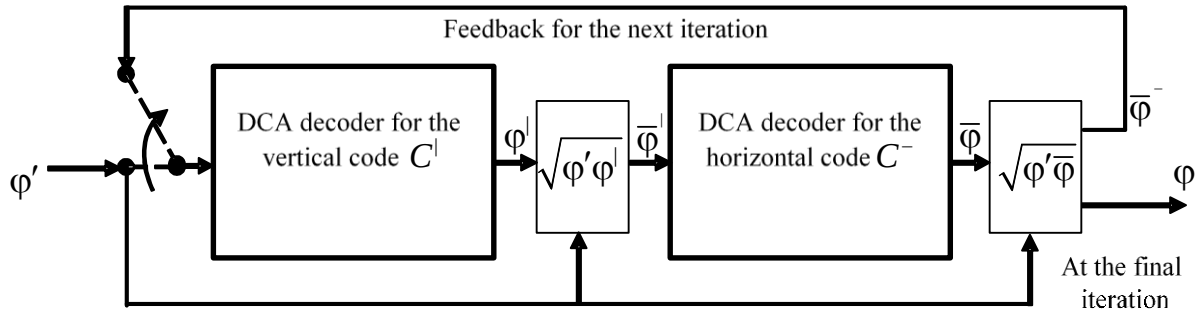


Fig. 2. Soft-decision decoding of dual codes for product codes

The decoder receives the signal matrix $y=\left[y_{uv},1\le u\le n_1,1\le v\le n_2\right]$ and calculate a matrix of the likelihood ratio for each corresponding bit node $\varphi'=\left[\varphi'_{uv},1\le u\le n_1,1\le v\le n_2\right]$, $\varphi'_{uv}=\exp(-2y_{uv}/\sigma^2)$. Set thematrix φ' as input of the proposed decoding algorithm (the Soft Decision Decoding of Dual Codes iterative Algorithm for Product Codes (SDDDCA-PC)). A decoder for product codes consists of two decoders: vertical $C^{|}$ and horizontal C^{-} in series. The SDDDCA-PC works as follows:

First, vertical (horizontal) decoder receives input message which is a matrix φ' and implement first iteration.

Step 1. Recalculate all values in each vertical (horizontal) in a matrix φ' to produce matrix $\varphi^{|}(n_2,n_1)$, *w*ith the corresponding value of bit node number *m* in any vertical (horizontal):

$$\varphi^{|}_m=\frac{\sum_{j=1}^{2^{n_2-k_2}}\prod_{l=1}^{n_2}\left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c_{2\,jl}^*}-\sum_{j=1}^{2^{n_2-k_2}}\prod_{l=1}^{n_2}\left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c_{2\,jl}^*\oplus\delta_{ml}}}{\sum_{j=1}^{2^{n_2-k_2}}\prod_{l=1}^{n_2}\left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c_{2\,jl}^*}+\sum_{j=1}^{2^{n_2-k_2}}\prod_{l=1}^{n_2}\left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c_{2\,jl}^*\oplus\delta_{ml}}}. \quad (9)$$

In which, $\oplus$ is the modul 2 addition; $\delta_{ml}=1$ if $m=l$ and $\delta_{ml}=0$ for other cases; $c_{a\,ij}'$ is the *l*-th bit of *j*-th codeword in the dual code $C'_a(n_a,r_a)$ of the original code $C_a(n_a,k_a)$, $a\in\{1,2\}$; $\varphi'_m=P\left(y_m|1\right)/P\left(y_m|0\right)$.

At the beginning of the vertical (horizontal) decoder, set the value of geometric average of matrix $\varphi^{|}$ and the initial input a message matrix φ', as the input of next horizontal (vertical) decoding step.

Step 2. The horizontal (vertical) directly receives the decoding message calculated in step 1, which is thematrix $\overline{\varphi}^{|}$:

$$\overline{\varphi}^{|} = \sqrt{\varphi^{|}\varphi'} \tag{10}$$

Similar to step 1, recalculae all values in each vertical (horizontal) in a matrix $\overline{\varphi}^{|}$ to produce the matrix φ^{-}, with value of bit node number *m* in each horizontal (vertical):

$$\varphi^{-}{}_{m} = \frac{\sum_{j=1}^{2^{n_1-k_1}} \prod_{l=1}^{n_1} \left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c^*_{1jl}} - \sum_{j=1}^{2^{n1-k_1}} \prod_{l=1}^{n_1} \left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c^*_{1jl}\oplus\delta_{ml}}}{\sum_{j=1}^{2^{n_1-k_1}} \prod_{l=1}^{n_1} \left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c^*_{1jl}} + \sum_{j=1}^{2^{n_1-k_1}} \prod_{l=1}^{n_1} \left(\frac{1-\varphi'_l}{1+\varphi'_l}\right)^{c^*_{1jl}\oplus\delta_{ml}}}. \tag{11}$$

in which, c^*_{1jl} is the *l*-th bit of *j*-th codeword in the dual code $C_1^*(n_1, r_1)$ of the original code $C_1(n_1, r_1)$. At the output of the second decoder, recalculatethe value of geometric average of matrix φ^{-}_{m} and the initial input message matrix φ' using equation (10) to get the matrix φ^{-}_{m} as the input for the next iteration. The decoding process is carried on until the final iteration.

Step 3. Decide the output codeword based on a matrix φ taken from the final iteration.

$$\begin{cases} C_{ij} = 1 \text{ when } \varphi_{ij} \geq 1 \\ C_{ij} = 0 \text{ in other cases,} \end{cases}$$

where C_{ij} is corresponding bit node in the product codeword $(1 \leq j \leq n_1, 1 \leq i \leq n_2)$.

Evaluation of performance of proposed decoding algorithm

The performance of the proposed algorithm for AWGN channel using Monte-Carlo simulation is conducted. The proposed algorithmis applied toproduct code, consistuted of Hamming codes (7,4); (15,11); (31,26), results are received after only two iteration as in Fig. 3.

Simulation results show that when using the SDDDCA-PC for product code, the signal to noise ratio required is only about 4,1 dB to achieve $BER = 10^{-6}$ with high coding ratio (0,7034). It is not necessary to increase number of iteration because this is optimal algorithms, so two iteration can provide reasonable decoding performance. The results also show that the product code with the longer distance of component codes and the higher code rate produces better the error control capacity. In order to evaluate the practical application for product codes when using the SDDDCA-PC decoding, we shall estimate number of calculations to be used in this algorithm. Table shows the number of calculations required for decoding one bit node through 2 iteration of the proposed algorithm. Evidently, the SDDDCA-PC has acceptable complexity which is linear function. Thus, this algorithm is very suitable with product codes whose component codes have small redundancy.

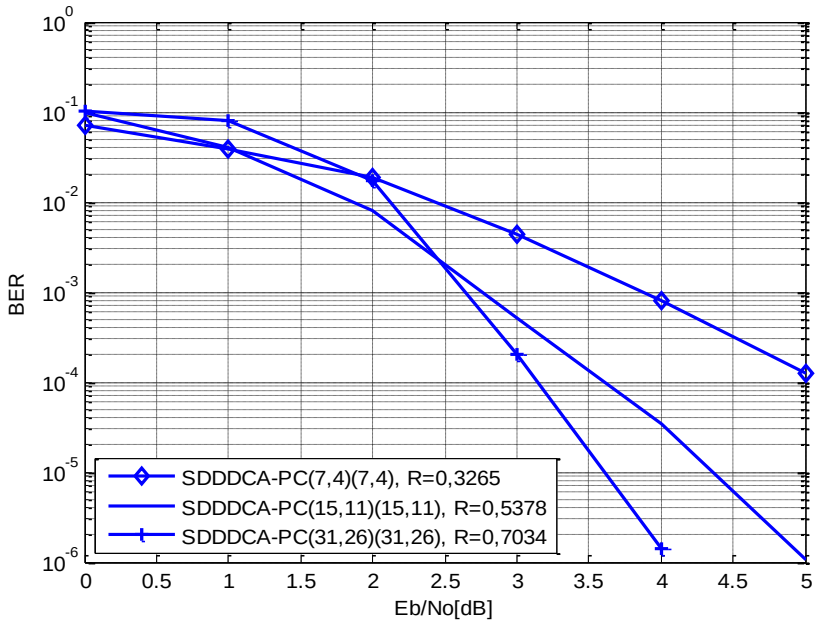


Fig. 3. Decoding performance of SDDDCA-PC for product code, consistuted of Hamming codes

Complexity of SDDDCA-PC

Algorithm	Number of multiplication	Number of addition
SDDDCA-PC	$2((n_1-1)2^{(r_1)}+(n_2-1)2^{(r_2)}+2)$	$2(2^{(r_1)}+2^{(r_2)}+2)$

Conclusion

In this article an algorithm for decoding product codes based on soft-decision decoding of component codes using dual code of the original code is proposed. The proposed decoding algorithm is based on strictly theoretical analysis and the Monte-Carlo simulation. Simulation results show that the algorithm is effective for product codes with component codes are codes with high coding rate. The proposed algorithm is applicable to modern communication systemswhereminimum processing latency is required.

References

1. Shannon C.E. // Bell Syst. Tech. J. 1948. Vol. 27. P. 379–423.
2. Morelos-Zaragoza R.H. The art of error correcting coding. John Wiley and Sons Inc., 2006.
3. Elias P. // IEEE Transactions on Information Theory. 1954. Vol. 4. P. 29–37.
4. Wolf J.K. // IEEE Transactions on Information Theory. 1978. Vol. 24. P. 76–80.
5. Hagenauer J., Offer E., Papke L. // IEEE Transactions on Information Theory. 1996. Vol. 42. P. 429–445.
6. Nickl H., Hagenauer J., Burkert F. // Proceedings of IEEE International Symposium on Information Theory. 1997.
7. Lim J.H., Lee J.H., Shin M.S., Han Cho G., Song Y.J. // Materials of International Conference on Control and Automation (CA). 2015. P. 13–16.
8. Son J., Kong J.J., Yang K. // Journal of Communications and Networks. 2018. Vol. 20. P. 345–353.
9. Pyndiah R.M. // IEEE Transactions on Communications. 1994. Vol. 46. P. 1003–1010.
10. MacWilliams F.J., Sloane N.J.A. The Theory of Error-Correcting Codes. New York (USA), 1981.

УДК 621.391

КОРРЕКЦИЯ ТРЕХКРАТНЫХ ОШИБОК БЧХ-КОДАМИ МЕТОДОМ СЖАТИЯ НОРМ СИНДРОМОВ

А.В. КУРИЛОВИЧ, А.Н. ПРИГОН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Рассмотрен метод коррекции трехкратных ошибок БЧХ-кодами, основанный на теории норм синдромов и сжатии множества ошибок корректируемой совокупности путем трансформации их в ошибки большей кратности. Это позволяет сократить число селектируемых комбинаций.

Ключевые слова: синдром ошибки, норма синдрома, БЧХ-код, декодер.

Введение

Синдромные методы коррекции ошибок эффективно реализованы в декодерах многих линейных кодов (Хемминга, БЧХ, РС и других). Однако их применение чаще всего ограничено кратностью $t \le 2$ корректируемых ошибок [1, 2]. Данное ограничение обусловлено «проблемой селектора» [3]: с ростом кратности t и длины кода n количество обрабатываемых синдромов и ошибок лавинообразно растет. Конструктивным решением «проблемы селектора» явилось применение норм синдромов [4] – нового параметра в помехоустойчивом кодировании информации, вычисляемого через компоненты синдрома, не зависящего от группы Г циклических сдвигов, позволяющего селектировать не отдельные ошибки, а их классы – Г-орбиты. Это в n раз сокращает количество селектируемых декодером комбинаций. Еще в $\log_2 n$ раз сокращается количество селектируемых комбинаций, применением циклотомических подстановок. Дальнейшим шагом в данном направлении является сжатие ошибок весом ω в специальный класс ошибок $M_{0,\omega+1}$, имеющих больший вес – $\omega+1$ и первую компоненту синдрома $s_1 = 0$. Ниже приведены описание и схема декодера для коррекции тройных ошибок БЧХ-кодом C_7 с проверочной матрицей $H = \left(\alpha^i, \alpha^{3i}, \alpha^{5i}\right)$, $0 \le i \le n-1 = 2^m - 2$ для примитивного элемента α поля Галуа $GF\left(2^m\right)$.

Структурная схема декодера, реализующего метод сжатия норм синдромов

Сжатию подвергаются векторы-ошибки $\bar{e}$ из множества $M_{\gamma,\omega}$, где $\gamma = s_1$ – первая компонента синдрома $S\left(\bar{e}\right)$, причем $\gamma \ne 0$. Они образуют полные Г-орбиты [4]. В каждой из этих Г-орбит содержится единственный вектор $\bar{f}$ с $s_1 = 1$. Каждый вектор Г-орбиты можно соответствующим циклическим сдвигом σ^{n-q} преобразовать в вектор $\bar{f}$. В [5] показано, что у векторов $\bar{f}$ весом 2, 3 первая координата равна нулю. Прибавив к $\bar{f}$ вектор $\bar{1} = (100...0)$, получим вектор $\bar{g} = \bar{f} + \bar{1}$ весом $\omega+1$ с $s_1 = 0$ и первой координатой, равной 1. Это множество обозначаем через $M^1_{0,\omega+1}$. Синдром ошибок $S\left(\bar{g}\right) = (0, N_1 + 1, N_2 + 1)$, где $N_1 = s_2 / s_1^3$, $N_2 = s_3 / s_1^5$ – координаты нормы синдрома $N\left(S\left(\bar{e}\right)\right)$, при этом $N\left(S\left(\bar{g}\right)\right) = (\infty, \infty, N_3)$, где $N_3 = s_3^3 / s_2^5$.

Реальной селекции подвергаются векторы $\bar{g}$, а точнее их синдромы и нормы синдромов. Векторов $\bar{g}$ весом 3, 4 в 2^{m-1} меньше, чем векторов $\bar{e}$ весом 2, 3. Векторы $\bar{g}$ принадлежат Г-орбитам, полным, если m – число простое.

Известно, что каждое значение нормы принимает в точности n различных синдромов [4]. Расчет показывает, что множество $M_{0,\omega}$ содержит Г-орбиты с одинаковым значением N_3 и, следовательно, с одинаковыми спектрами синдромов. Но при этом происходит максимально возможно плотная упаковка (и соответственно максимальное сжатие множества селектируемых норм) векторов $\bar{g}$: найдется в точности n различных векторов $\bar{g}$ с данным значением N_3 и с попарно различными синдромами (см. табл. 1).

Таблица 1. **Г-орбиты ошибок (одна весом 3 и 3 весом 4) из $M_{0,3}$ и $M_{0,4}$ с показателями ($\deg s_2$, $\deg s_3$) в (15,3,7) – БЧХ-коде С с нормой $\bar{N} = (\infty, \infty, \alpha^5)$**

$M_{0,3}$	$M_{0,4}$	$M_{0,4}$	$M_{0,4}$	($\deg s_2$, $\deg s_3$)
(1,12,13)	(3,8,9,10)	(4,6,11,15)	(2,5,7,14)	(8,10)
(2,13,14)	(4,9,10,11)	(1,5,7,12)	(3,6,8,15)	(11,0)
(3,14,15)	(5,10,11,12)	(2,6,8,13)	(1,4,7,9)	(14,5)
(1,4,15)	(6,11,12,13)	(3,7,9,14)	(2,5,8,10)	(2,10)
(1,2,5)	(7,12,13,14)	(4,8,10,15)	(3,6,9,11)	(5,0)
(2,3,6)	(8,13,14,15)	(1,5,9,11)	(4,7,10,12)	(8,5)
(3,4,7)	(1,9,14,15)	(2,6,10,12)	(5,8,11,13)	(11,10)
(4,5,8)	(1,2,10,15)	(3,7,11,13)	(6,9,12,14)	(14,0)
(5,6,9)	(1,2,3,11)	(4,8,12,14)	(7,10,13,15)	(5,5)
(6,7,10)	(2,3,4,12)	(5,9,13,15)	(1,8,11,14)	(2,0)
(7,8,11)	(3,4,5,13)	(1,6,10,14)	(2,9,12,15)	(14,10)
(8,9,12)	(4,5,6,14)	(2,7,11,15)	(1,3,10,13)	(11,5)
(9,10,13)	(5,6,7,15)	(1,3,8,12)	(2,4,11,14)	(8,0)
(10,11,14)	(1,6,7,8)	(2,4,9,13)	(3,5,12,15)	(5,10)
(11,12,15)	(2,7,8,9)	(3,5,10,14)	(1,4,6,13)	(2,5)

Вычислив $S(\bar{g})$ и N_3, однозначно определяем вектор $\bar{g}$ и, следовательно, вектор $\bar{f}$, циклический сдвиг $\sigma^q(\bar{f}) = \bar{e}$ есть истинное значение искомого вектора ошибок. Предложенная схема коррекции тройных ошибок реализована декодером (рис. 1). Здесь БВС – блок вычисления синдрома $S(\bar{x}) = S(\bar{e})$, БАС – блок анализа синдрома, БВОВО – блок вычисления образующего вектора ошибок $\bar{f}$, БВТВО – блок вычисления текущего вектора ошибок $\bar{e}$, БИ – блок инвертации (вычисляет сумму $\bar{x} + \bar{e}$).

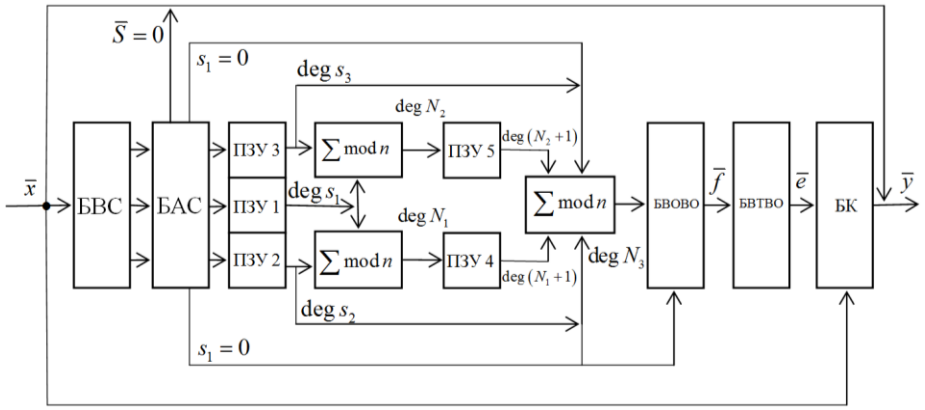


Рис. 1. Структурная схема декодера, реализующая метод сжатия с помощью ПЗУ и сумматоров по модулю *n* для двоичного БЧХ-кода с минимальным расстоянием 7

Поясним работу декодера. Если $S = 0$, то принятое сообщение $\bar{x}$ не содержит ошибок и поступает на выход декодера.

Если $s_1 \neq 0$, то ПЗУ 1, 2 и 3 вычисляют показатели компонент s_1, s_2, s_3 синдрома $S(\bar{x}) = (s_1, s_2, s_3)$ соответственно, которые поступают на сумматоры 1 и 2 по модулю n. Первый из них вычисляет $\deg N_1 = \deg s_2 - 3\deg s_1$, $\deg N_2 = \deg s_3 - 5\deg s_1$. На выходах ПЗУ 4 и 5 значения показателей $\deg(N_1 + 1)$ и $\deg(N_2 + 1)$ соответственно, которые поступают на сумматор 3 по модулю n для вычисления $\deg N_3$ – третьей координаты нормы синдрома вектора $\bar{g}$. Может оказаться, что $N_1 = N_2 = 1$. Тогда $\bar{e}$ – одиночная ошибка на позиции $\deg s_1 + 1$. Если $s_1 = 0$, то показатели $\deg s_2$ и $\deg s_3$ непосредственно поступают на третий сумматор для вычисления $\deg N_3 = 3\deg s_3 - 5\deg s_2$ вектора ошибки $\bar{e}$ заведомо принадлежащего множеству $M_{0,3}$.

Заключение

Норменные декодеры на ПЗУ при $t = 2$ примерно в 20 раз проще по аппаратурным затратам известных синдромных декодеров и в 300 раз проще декодеров, аппаратно решающих уравнения в полях Галуа. Предложенный декодер тройных ошибок имеет сложность, сопоставимую со сложностью норменного декодера при $t = 2$, а количество селектируемых норм и векторов существенно меньшее. Так же возможна реализация метода сжатия на ПЛИС.

TRIPLE ERRORS CORRECTION BY BCH-CODES BY THE COMPRESSION METHOD OF SYNDROMES NORMS

A.V. KURYLOVICH, A.N. PRIGON

Abstract. The correction method of triple errors by BCH-codes is considered. It based on the theory of syndromes norms and compression correcting set errors by transforming them into errors of greater weight. It allows to reduce the number of selectable combinations.

Keywords: error syndrome, syndrome norm, BCH-code, decoder.

Список литературы

1. Сагалович Ю.Л. // Автоматика и телемеханика. АиТ. 1991. С. 135–145.
2. Муттер В.М., Петров Г.А., Маринкин В.И., Степанов В.С. Микропроцессорные кодеры и декодеры. М.: Радио и связь, 1991.
3. Мак-Вильямс Ф.Дж. Перестановочное декодирование систематических кодов. Кибернетический сборник. Новая серия. М.: Связь, 1965.
4. Конопелько В.К, Липницкий В.А. Теория норм синдромов и перестановочное декодирование помехоустойчивых кодов. Минск.: БГУИР, 2000.
5. Конопелько В.К., Липницкий В.А., Курилович А.В. // Докл. 5-ой МНТК «Цифровая обработка сигналов и ее применения»: Т. 1. Москва. 12–14 марта. 2003 г. С. 146–149.

УДК 621.392

IMPLEMENTATION OF INTELLIGENT BIOMEDICAL DIAGNOSTICS BASED ON THE FUZZY INFERENCE SYSTEM

Al SABEEH AMJAD KARIM, NGUEN HONG KUAN, M.Yu. HOMENOK

Belarussian state university of informatics and radioelectronics, Republic of Belarus

Submitted 15 March 2019

Abstract. Intelligent biomedical diagnosis system is one of the most active research areas nowadays. Use of artificial intelligence and computer technologies in diagnosis and treatment of illnesses or training of fitness has highly increased. Heart disease diagnosis is a challenging task which can offer automated prediction about the heart disease of patient so that further treatment can be made easy. In this article the fuzzy logic based heart disease diagnosis system is analyzed on base sixth risk that affect on heart disease risk.

Keywords: fuzzy logic, fuzzy expert systems, fuzzy inference system, making decision.

Introduction

In several areas of biomedical domain, including prediction of the effectiveness of surgical procedures, medical tests and the discovery of relationships among clinical and diagnosis data, data mining techniques have been applied. Modern-day medical diagnosis is a very composite process, entailing precise patient data, a philosophical understanding of the medical literature and many years of clinical experience. The health care data which, unfortunately, are not «mined» to discover hidden information for effective decision-making are collected in a huge amount by the health care industry. Use of artificial intelligence or computer technology in the fields of medicine area diagnosis and treatment of illnesses has highly increased. The biomedical field has a challenging field because of very high complexity and uncertainty. Therefore, the use of intelligent systems such as fuzzy logic has been developed. Because of the many and uncertain risk factors in the heart disease risks, sometimes heart disease diagnosis is hard for experts. In the other word, there exists no strict boundary between what is Healthy and what is diseased, thus distinguish is uncertain and vague.

There are huge data management tools available within health care systems, but analysis tools are not sufficient to discover hidden relationships amongst the data. Most of the medical information is vague, imprecise and uncertain. Medical diagnosis is a complicated task that requires operating accurately and efficiently.

Fuzzy set theory and fuzzy logic have a number of characteristics that make them highly suitable for modeling uncertain information upon which medical concept forming, patient state interpretation, and diagnostic as well as therapeutic decision making is usually based. Firstly, medical entities such as symptoms, signs, test results, diseases and diagnoses, therapeutic and prognostic information can be defined as fuzzy sets. The inherent vagueness of these entities will thus be conserved. Secondly, fuzzy logic offers reasoning methods capable of drawing strict as well as approximate inferences. Medicine demands this broad range of possibilities because the body of medical theory includes definitional, causal, statistical, and heuristic knowledge. Practical medicine even has to accept incomplete medical theories where only vague and uncertain empirical information guides the necessary medical procedures. Finally, fuzzy automata maybe used as high level patient monitoring devices with real time access to medical information systems.

The advantages of fuzzy logic are its simplicity, flexibility of combining conventional control techniques, ability to model nonlinear functions and imprecise information, use of empirical knowledge and dependency on heuristics.

A fuzzy logic system is mainly comprised of four components: fuzzifier, defuzzifier, fuzzy rule base and fuzzy inference engine. these components are arranged as follows in any fuzzy logic system, (Fig. 1).

Fuzzification is the first process that takes place in the FLS. A numeric or crisp input value is given to the fuzzifier. The crisp input value is required to be converted to the corresponding fuzzy value as the rules for determining the result, are defined for fuzzy inputs. This task is performed by the fuzzifier and then the fuzzy input values are supplied to the fuzzy inference engine, which is responsible for computing the set of outputs based on the IF-THEN rules defined in the fuzzy rule base.

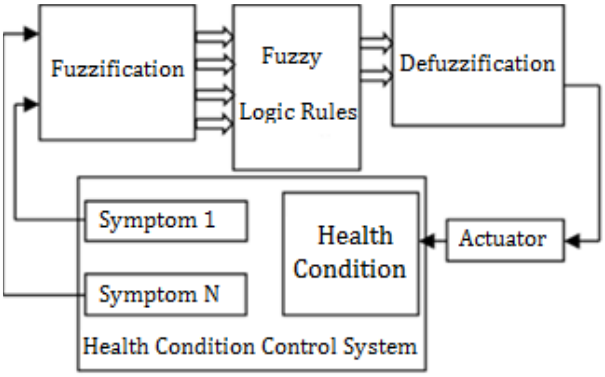


Fig. 1. Block diagram of health condition using fuzzy logic system

Usually, when more than one inputs are required, AND operator is used to combine them. The last process in the fuzzy logic system is defuzzification. It converts the fuzzy output values into their corresponding crisp values. There are different methods for fuzzification and defuzzification. Some widely used fuzzifiers are Singleton fuzzifier, Gaussian fuzzifier and trapezoidal or triangle fuzzifier. Singleton fuzzifier is the simplest fuzzifier which basically assigns a precise value to the given input and hence no fuzziness is introduced by fuzzification in this case. Gaussian and Triangular fuzzifiers are used to suppress the noise in the given inputs. Examples of defuzzifiers are maximum defuzzifier, mean of maxima defuzzifier, centroid defuzzifier, height defuzzifier, modified height defuzzifier, center of sets and center of sums.

Dataset for design of fuzzy system

Dataset for the designed system is collected from various Hospitals and by consulting the expert in the field of heart disease [1, 2]. The purpose of this dataset is to diagnose the presence or absence of heart disease given the results of various medical tests carried out on a patient. This system uses 6 attributes for input and 1 attribute for output.

The Input attributes are chest pain type, blood pressure, cholesterol, blood sugar, maximum heart rate and old peak (ST depression induced by exercise relative to rest). The output field refers to the presence of heart disease in the patient. It is integer value from 0 (no presence) to 1; increasing value shows increasing heart disease risk. Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise. In contrast with binary sets having binary logic, also known as crisp logic, the fuzzy logic variables may have a membership value of only 0 or 1. Firstly there is a need to make a dataset to decide the range of parameters on which the heart diseases are depending.

Chest pain. In input attribute Chest pain, we choose five different membership functions which are very low, low, moderate, high and very high. The range of this attribute is given in Table 1 and on Fig. 2.

Table 1. **Classification of chest pain**

Input Field	Range	Fuzzy set
CHEST PAIN	1-3	Typical angina
	3-5	Atypical angina
	5-7	Non-anginal
	7-9	Asymptomatic

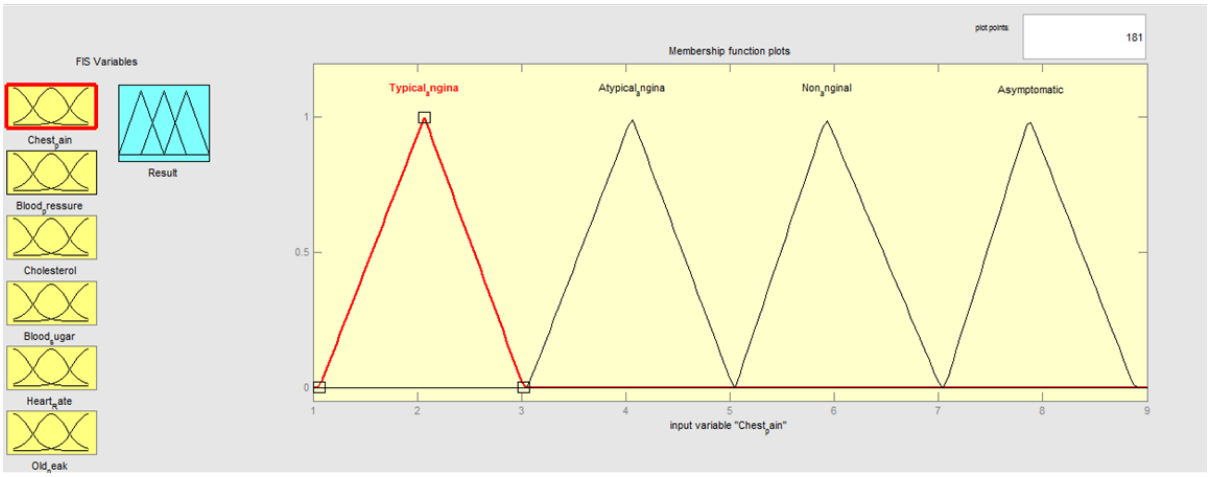


Fig. 2. Membership functions of chest pain

Blood pressure. We choose five different membership functions for input attribute blood pressure and are named as very low, low medium, high and very high. This can also be shown in Table 2 and on Fig. 3.

Table 2. **Classification of blood pressure**

Input Field	Range	Fuzzy set
BLOOD PRESSURE	<117 mmHg	low
	129–147 mmHg	medium
	139–160 mmHg	high
	>173 mmHg	very high

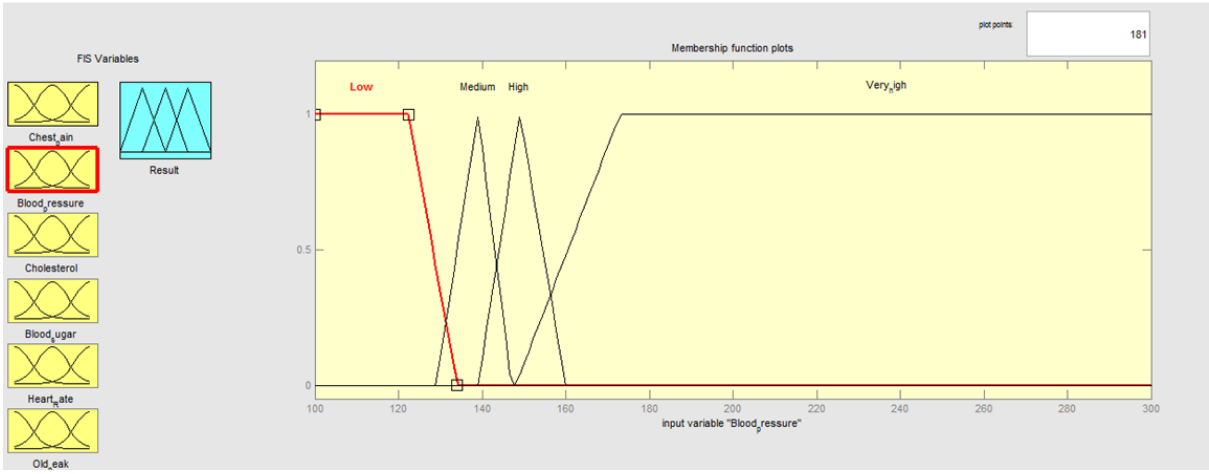


Fig. 3. Membership functions of blood pressure

Cholesterol. Cholesterol has salient affect on the result and can change it easily. Authors chose five different membership functions for input attribute cholesterol. The range and membership function of this attribute is shown in Table 3 and on Fig. 4.

Table 3. **Classification of cholesterol**

Input Field	Range	Fuzzy set
CHOLOSTEROL	<136 mg/dl	low
	188–250 mg/dl	medium
	217–307 mg/dl	high
	>342 mg/dl	very high

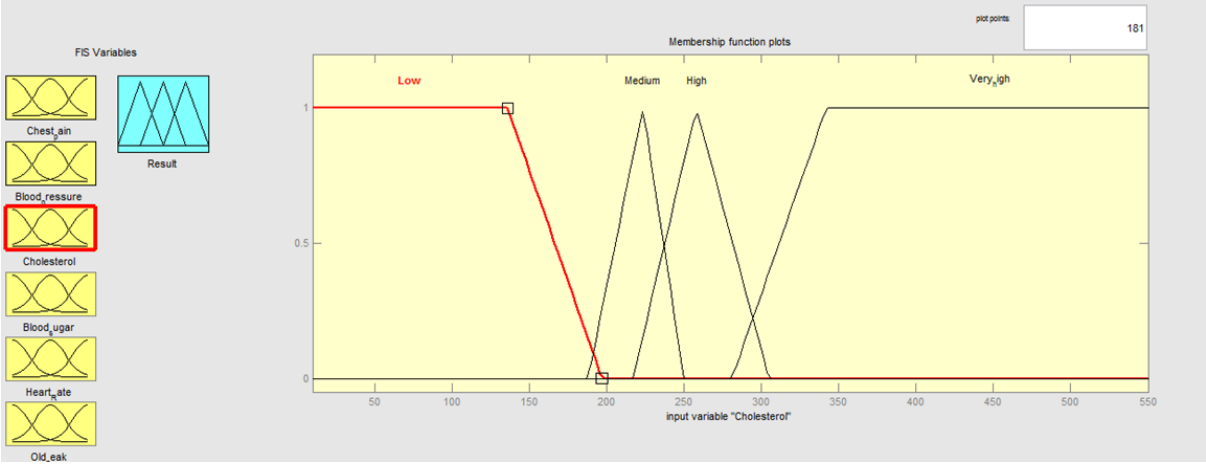


Fig. 4. Membership functions of cholesterol

Blood sugar. Blood sugar field is one of the most important factors in the diagnostics of heart disease. The input attribute blood sugar, five different membership functions were used which is given as very low, low, medium, high and very high. The range and the membership function of this attribute is given in Table 4 and on Fig. 5.

Table 4. **Classification of blood sugar**

Input Field	Range	Fuzzy set
BLOOD SUGAR	60–240 mg/dl	low
	>244	high

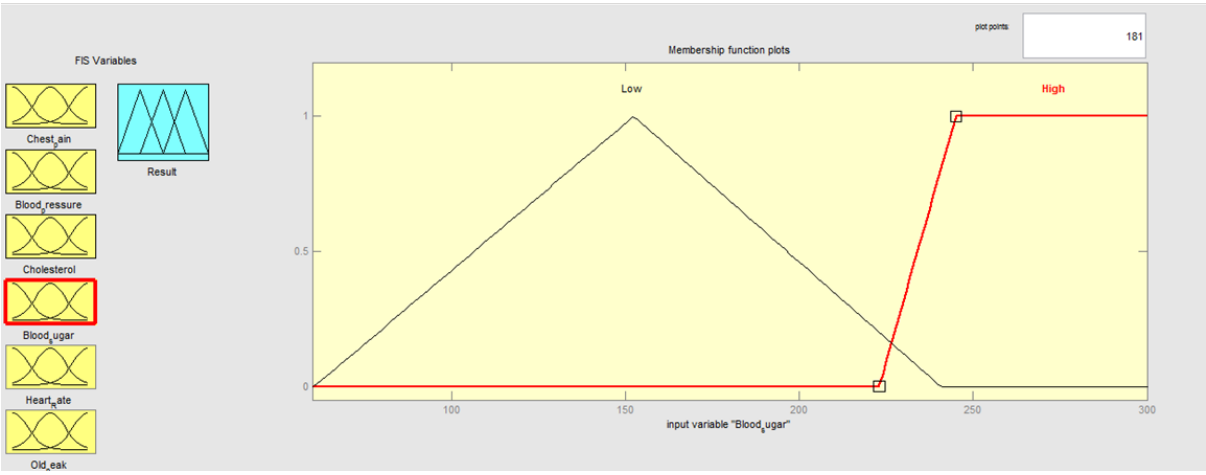


Fig. 5. Membership functions of blood sugar

Heart rate. Heart rate has five different membership functions, i.e. very low, low, medium, high and very high. The range and membership function of this attribute is given in Table 5 and on Fig. 6.

Table 5. **Classification of heart rate**

Input Field	Range	Fuzzy set
HEART RATE	<80 bpm	low
	111–194 bpm	medium
	>220 bpm	high

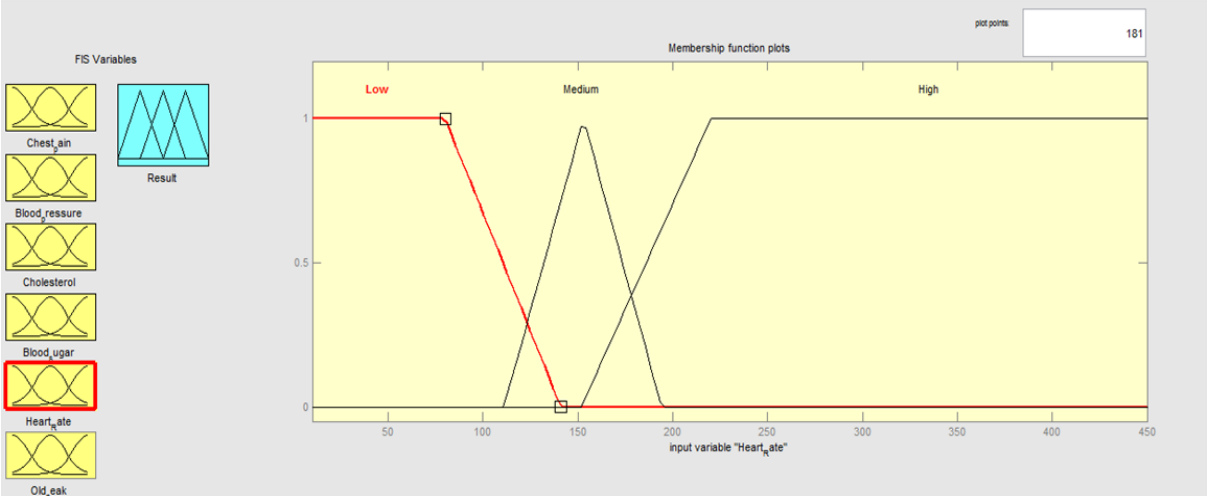


Fig. 6. Membership functions of heart rate

Old peak. This input attribute means ST depression induced by exercise relative to rest. Old peak field has 5 fuzzy sets (very low, low, medium, terrible and risk). These fuzzy sets have been shown in Table 6 with their ranges and on Fig. 7.

Table 6. **Classification of old peak**

Input Field	Range	Fuzzy set
OLD PEAK	<1	low
	1,5–4,2	risk
	>4	terrible

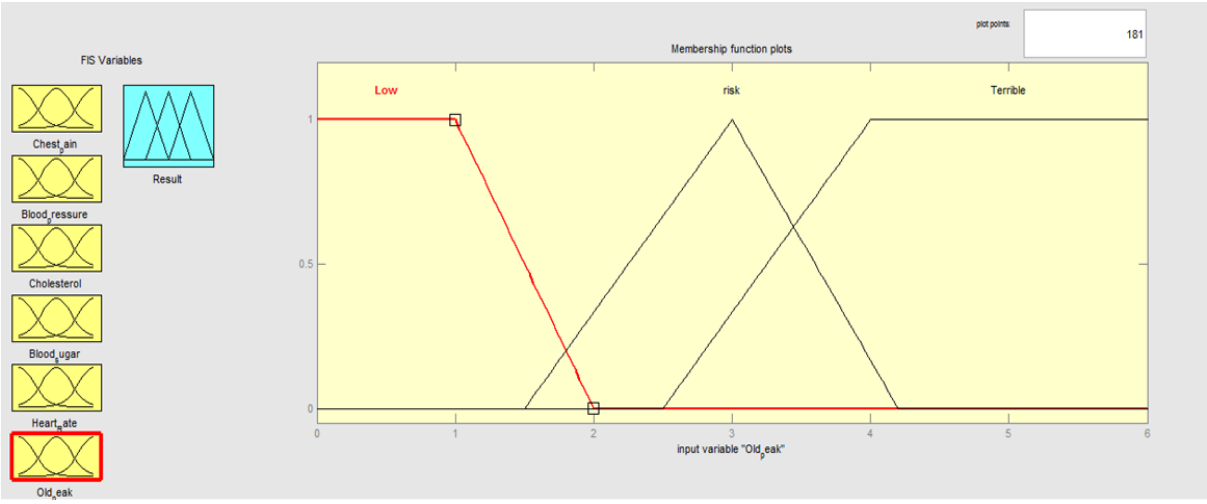


Fig. 7. Membership functions of old peak

System Testing

The system uses Mamdani inference mechanism. The result shows the presence of heart disease risk in patient. It has value from 0 (no presence) to 1. There are 4 Fuzzy sets: healthy, sick1 – low risk,

sick2 – moderate risk, sick3 – high risk. Inference engine of this system is designed with the help of the expert and contains 18 dependent rules that uses forward chaining inference mechanism to accurately diagnose the heart disease risk. Centroid technique is used for defuzzification. The results of system testing have been shown in Table 7 and on Fig. 8–11.

Table 7. **Fuzzy sets range of output variable**

Input Field	Range	Fuzzy set
OUTPUT VARIABLE	<1,5	healthy
	2,5–4,5	sick1
	5,5–7,5	sick2
	>8,5	sick3

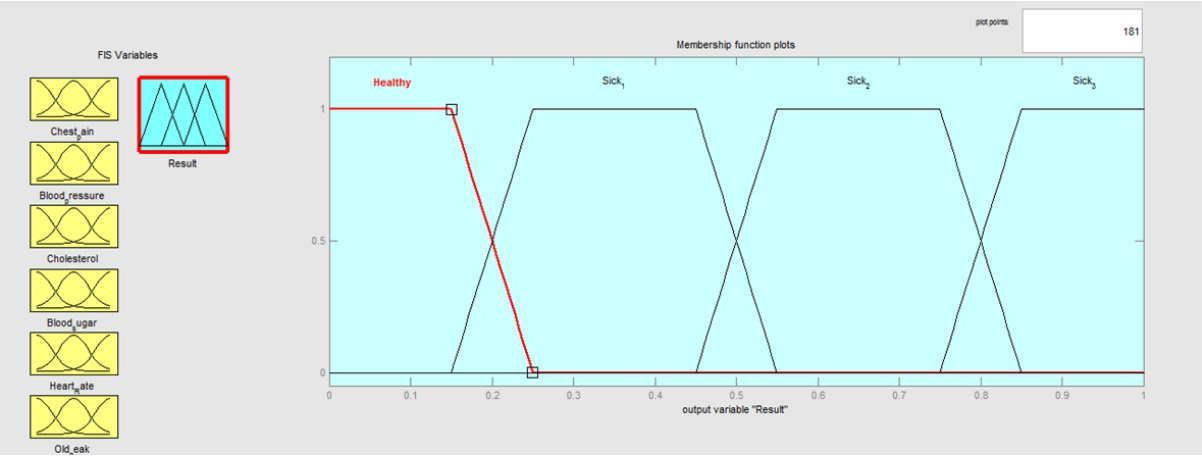


Fig. 8. Membership functions of result

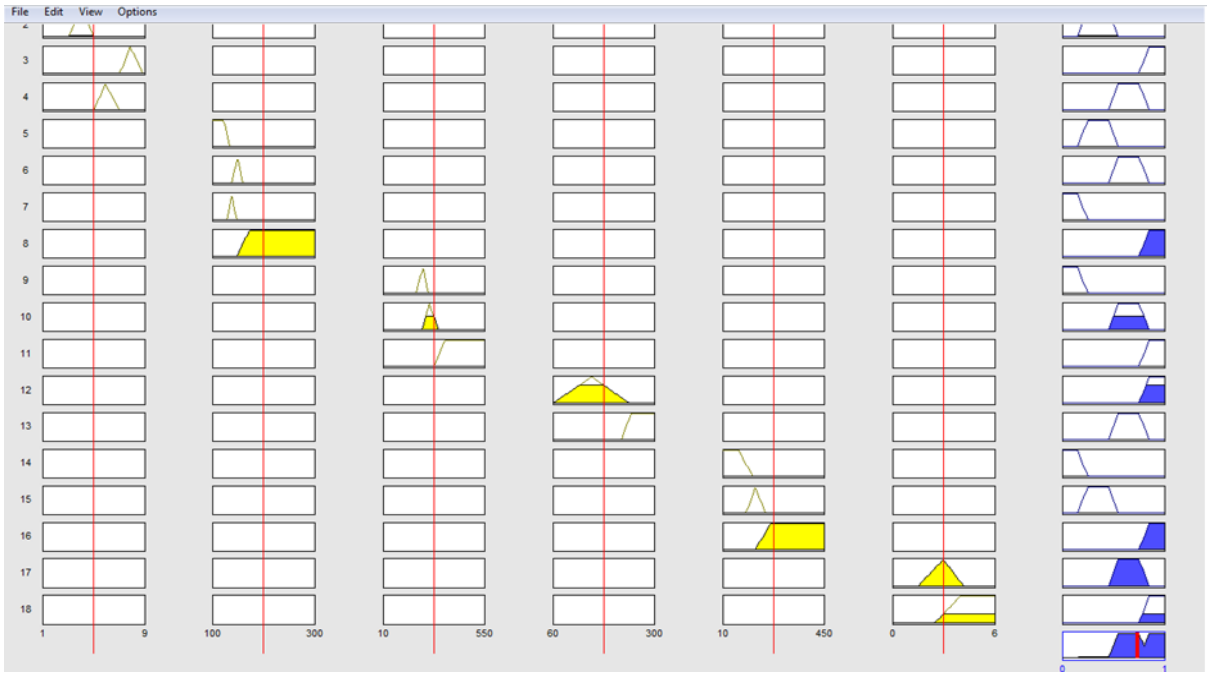


Fig. 9. Result of fuzzy system

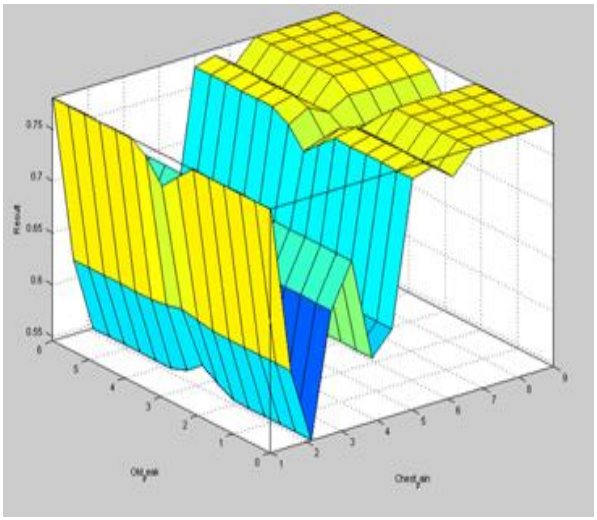


Fig. 10. Surface viewer of chest pain and old peak

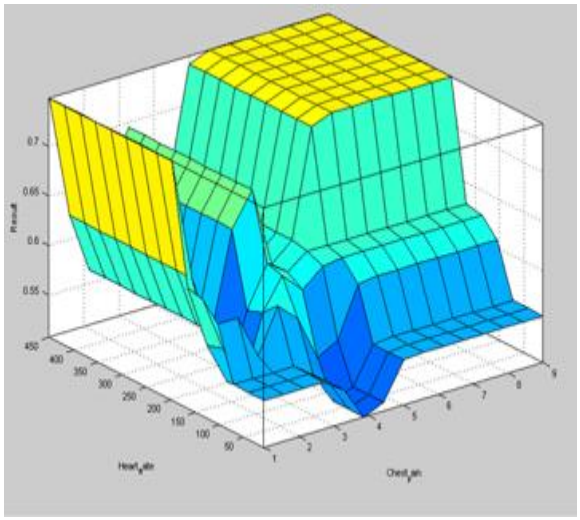


Fig. 11. Surface viewer of chest pain and heart rate

Conclusion

Analysis performed in fuzzy logic tools [3] using of the fuzzy inference system for medical diagnosis shows what this technology may be used in private life and to improve decision making about state of patients in hospitals or during the process of transporting the patient. Experimental results according to numerous studies showed that this system did quite better than non-expert and about 92 % as a well as the expert did [1].

References

1. Multicare Health System: Using a Modified Early Warning System (MEWS) to Improve Patient Safety HIMSS Innovation Community November 2, 2012.
2. Gawande P.S. // National Conference on Innovative Trends in Science and Engineering (NC-ITSE'16). Vol. 7. P. 31–36.
3. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику [Electronic resource]. URL: https://matlab.exponenta.ru/ (date of access: 10.03.2019).

УДК 004.052.32

ФУНКЦИОНАЛЬНОЕ ТЕСТИРОВАНИЕ СЕТЕВЫХ КОМПОНЕНТОВ СИСТЕМЫ «УМНЫЙ ДОМ»

М.А. АЛИСЕЕНКО, Б.В. НИКУЛЬШИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Рассмотрены структура и функционирование системы «Умный дом». Описаны компоненты ее устройств. Разработан подход тестирования сетевых компонентов системы «Умный дом».

Ключевые слова: «Умный дом», сетевые компоненты, функциональное тестирование.

Введение

Системы «Умный дом» становятся все более распространенными и доступными благодаря тому, что помогают человеку повысить уровень комфорта, снизить затраты на электроснабжение, обеспечить контроль за внешними факторами. В настоящее время актуальной задачей является формирование комплексной стратегии тестирования компонентов умного дома, обеспечивающих необходимое качество предоставляемых пользователю услуг [1, 2]. Взаимодействие устройств системы осуществляется по сети, состояние которой может повлиять на производительность, качество работы устройств и инфраструктуры в целом [3]. Поэтому система должна быть протестирована для гарантии бесперебойной работы. Целью работы является создание подхода для тестирования сетевых компонентов системы «Умный дом», характеризующегося невысокими затратами. Для этого необходимо выделить функциональные блоки устройств, отвечающие за одни и те же управляющие команды, чтобы проверить их работоспособность. Если эти показатели проходят проверку, то можно судить о качественной работе компонентов в системе «Умный дом».

Схема базовых проверок компонентов системы «Умный дом»

Функциональное тестирование сетевых компонентов системы «Умный дом» предлагается разделить на три уровня:

– проверка физических параметров (частотный диапазон, мощность, дальность приемопередатчика);

– проверка отрабатывания команд силовыми блоками (программного обеспечения устройств);

– проверка API (корректность взаимодействия низко- и высокоуровневых компонентов).

Радиоприемные устройства беспроводной системы являются исполнительными силовыми блоками. Радиопередающие устройства – управляющие радиопульты. Для приема команд с радиопередающих устройств системы используется адаптер RX2164 (RX1164), для передачи команд – адаптер PC1132 или Ethernet-шлюз (контроллер). Устройства также разбиваются на типы в зависимости от выполняемых функций: управление освещением, отоплением, двигателем, пульты, датчики.

На первом уровне резонаторы на поверхностных акустических волнах (ПАВ-резонаторы) пультов и датчиков проверяются на тестовом стенде для выделения несущей. Блоки питания и силовые элементы подвергаются имитации частого включения и выключения из сети при различных температурных условиях, что может привести к нагреву и выходу из строя внутренних компонентов. Управляющие элементы проверяются на соответствие максимальной выходной характеристике мощности. Все типы устройств должны принимать команды на расстоянии, соответствующем техническим характеристикам.

На втором уровне функциональные компоненты тестируются на корректность отработки запрограммированных команд. Минимальный набор команд для силовых блоков включения (on), выключения (off), привязки (bind), отвязки (unbind/clear) возможно проверить с помощью универсального пульта. Каждый силовой блок имеет стандартный идентификатор (id) универсального пульта, благодаря чему не требует привязки пользовательского пульта (PR) или датчика (PT). Стандартная привязка должна подтвердиться на силовом блоке за счет нажатия сервисной кнопки, аналогичным образом производится отвязка, а затем очистка памяти устройства (рис. 1). Шесть операций проверки можно свести к трем.

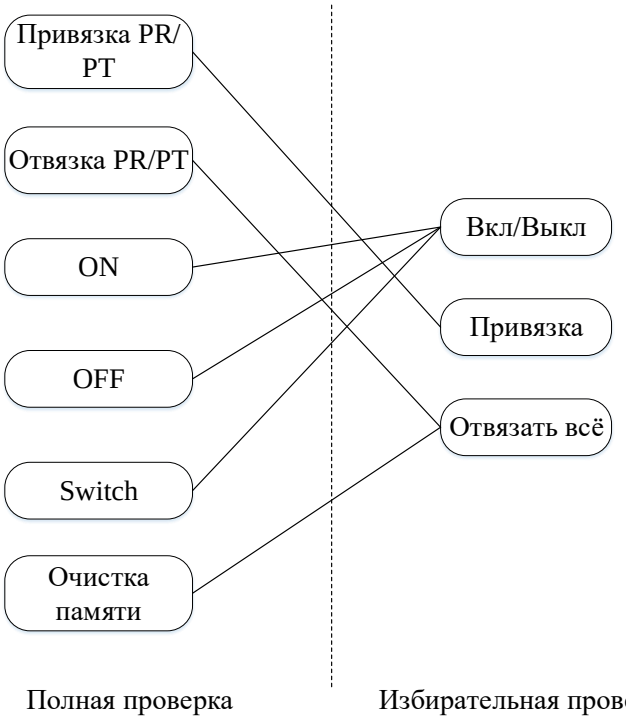


Рис.1. Проверка силовых блоков с помощью универсального пульта

Третий уровень подразумевает проверку управления устройствами посредством пользовательского программного обеспечения. Команды, поддерживаемые устройствами, можно разделить на подгруппы и посылать эти подгруппы на тестовый стенд. Это является более эффективным, чем поочередная отправка каждой из команд.

Схема проверки для функционального блока датчика температуры

Термостатный блок управления (Т) состоит из блока питания (БП), центрального процессора (ЦП), приемопередатчика с антенной, датчика температуры, связанного с нагрузкой (нагревательным элементом) и ЦП. Блок RF выделяет информационную составляющую от принятой команды управления. Полученная информация обрабатывается в блоке ЦП. Команда, направленная на взаимодействие с нагрузкой, формирует в ЦП управляющий импульс на силовой элемент (реле), если датчик температуры зарегистрирует температуру ниже заданного порога. Если принятой командой была команда изменения настроек, то она также обрабатывается блоком MCU, изменяя настройки силового блока, после чего полученные изменения сохраняются в энергонезависимую память MCU (SRAM). На каждую принятую по радиоканалу команду силовой блок должен дать команду ответа, содержащую в себе актуальное состояние нагрузки. Команда ответа формируется блоком MCU и передается на блок RF, где происходит ее модуляция и транслирование в радиоэфир. Функциональная схема термостатного блока управления представлена на рис. 2.

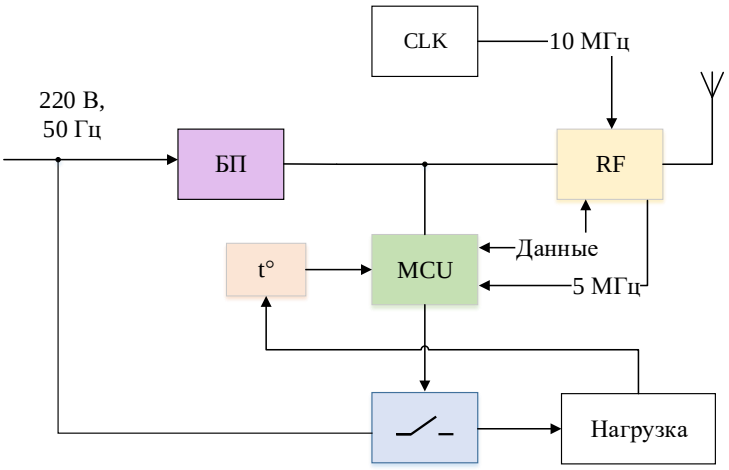


Рис. 2. Функциональная схема термостатного блока управления

Датчик температуры и влажности (РТ1) состоит из питающего элемента, центрального процессора, передатчика (ТХ) с антенной. ЦП обрабатывает данные с датчиков температуры ($t°$) и влажности (Ψ). Регулятором порога выступает подстроечный резистор. Если датчик зафиксировал величину ниже или выше порога, то блок ЦП формирует команду для отправки н привязанный силовой блок. Преобразованная команда отправляется блоком ТХ в радиоэфир. Функциональность датчика температуры (РТ2) аналогична (в нем отсутствует датчик влажности). Функциональная схема устройства контроля температуры и влажности показана на рис. 3.

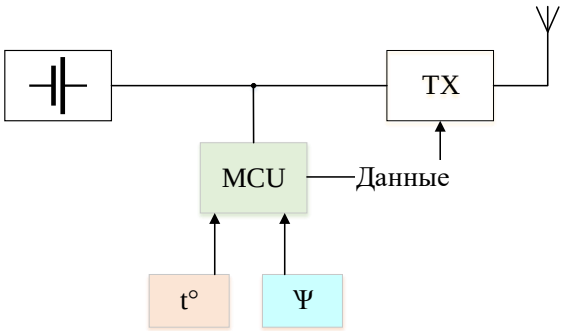


Рис. 3. Функциональная схема устройства контроля температуры и влажности

Для функциональной проверки группы команд, основанных на показаниях датчика температуры (и влажности), следует прибегнуть к тестовому соединению ТХ передатчика датчика температуры (и влажности) с ЦП блока термостата. Таковыми командами являются включение нагрузки (on), выключение нагрузки (off), переключение нагрузки (switch), передача данных о температуре/влажности (состоянии) элемента (sens). С помощью тестового стенда датчик температуры подвергается воздействию изменяющейся температурой, в результате чего датчик последовательно формирует четыре команды на ЦП блока термостата и ЦП датчика температуры и влажности. Далее должно произойти отрабатывание команд нагревающими элементами, связанных с датчиком. Затем необходимо проверить значение пришедшей температуры и индикатор состояния силового блока в пользовательском приложении. Таким образом, проверяется способность функционального компонента ЦП силового блока и датчиков формировать команды изменения состояния нагрузки и передачи данных о состоянии, способность приема команды блоком RF блока термостата, способность передачи блока ТХ датчиков.

Обычная проверка функционального элемента $t°$ в блоке термостата, датчике температуры, датчике температуры и влажности потребовала бы ожидания изменения значения температуры относительно порога на устройствах, чтобы нагревающий элемент включился или выключился, привязки и отвязки датчиков температуры (и влажности) к силовому блоку термостата

поочередно. Граф проверок команд блока *t*° для термостатного блока и датчиков представлен на рис. 4.

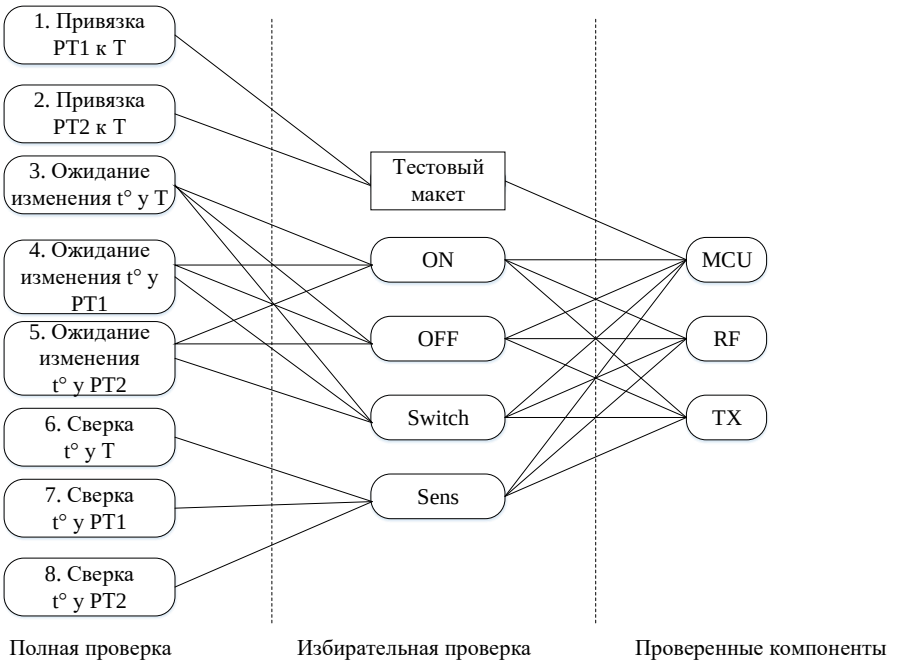


Рис. 4. Проверка функционального элемента контроля температуры

Из рис. 4 следует, что восемь операций для проверки ручным способом могут быть сведены к 4 командам, выполняемым на тестовом стенде и предназначенным для проверки работоспособности функциональных блоком MCU, RF, TX.

Заключение

Разработан подход к функциональному тестированию сетевых компонентов системы «Умный дом», позволяющий сократить количество производимых операций при ее ручной проверке и снизить за счет этого временные и экономические затраты на этапе тестирования.

FUNCTIONAL TESTING OF «SMART HOME» NETWORK COMPONENTS

M.A. ALISEYENKA, B.V. NIKULSHYN

Abstract. The structure and functioning of the «Smart home» system is considered. The system is described. The testing approach of its network components is elaborated.

Keywords: «Smart home», network component, functional testing.

Список литературы

1. Борисова М.В., Киричек Р.В. // Информационные технологии и телекоммуникации. 2018. Т. 6. № 2. С. 27–34.
2. Кулик В.А., Киричек Р.В., Бондарев А.Н. // Информационные технологии и телекоммуникации. 2015. № 2 (10) С. 106–114.
3. Долгушев Р.А., Киричек Р.В., Кучерявый А.Е. Обзор возможных видов и методов тестирования Интернет Вещей // Информационные технологии и телекоммуникации. 2016. Т. 4. № 2. С. 1–11.

УДК 621.391.1

УНИВЕРСАЛЬНЫЙ FSK/QAM МОДУЛЯТОР НА ПРОГРАММИРУЕМОЙ ЛОГИЧЕСКОЙ ИНТЕГРАЛЬНОЙ СХЕМЕ ALTERA CYCLONE V

А.Л. ХОМИНИЧ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Рассмотрена реализация на программируемых логических интегральных схемах (ПЛИС) семейства Intel/Altera Cyclone V цифрового модулятора, обеспечивающего формирование частотно-манипулированных и квадратурно модулированных радиосигналов с размерностью сигнального созвездия от 2 до 1024. Предложены варианты реализации модуляторов, обеспечивающие эффективное использование ресурсов ПЛИС.

Ключевые слова: манипуляция, модуляция, интерполяция, сигнальное созвездие, преобразование частоты.

Введение

Цифровые системы связи уже практически полностью вытеснили своих аналоговых предшественников. Наиболее используемым видом модуляции (если подходить строго, то манипуляции) является квадратурная амплитудная (QAM). Она применяется как в системах с одной несущей, так и в многочастотных, с ортогональным разделением каналов (OFDM), при этом размер сигнального созвездия составляет обычно от 4 до 1024, а в ряде случаев, в частности, в проводных и кабельных системах, встречаются созвездия с размером до 16384.

Частотная манипуляция (FSK) используется в настоящее время гораздо реже, прежде всего, из-за меньшей по сравнению с QAM помехоустойчивостью (при одинаковой спектральной эффективности). Количество ступеней девиации частоты редко превышает 16, хотя сведения о моделировании и теоретических исследованиях систем с большим размером созвездия также присутствуют. Преимуществом FSK является более простая реализация приемного тракта.

Теоретические аспекты формирования и применения FSK и QAM подробно описаны в учебной и научно-технической литературе [1, 2], поэтому не обсуждаются в настоящей статье. Основной целью приведенной работы являлась практическая реализация универсального FSK/QAM модулятора на ПЛИС Intel/Altera Cyclone V с использованием отладочных модулей Terasic DE10-Standard и GX starter Kit. Для подсистемы цифро-аналогового преобразования использована плата расширения Terasic THDB-ADA.

Общая структурная схема цифровой части радиопередающего тракта

При традиционном построении цифровой передающий тракт (рис. 1) должен включать в себя входной интерфейс, выполняющий прием и адаптацию входного потока данных, блок помехоустойчивого кодирования (БПК), модуляторы – в данном случае частотный (ЧМГ) и квадратурный (КАМ) – цифро-аналоговый преобразователь (ЦАП). Также в любом случае необходим генератор тактовых сигналов (ГТС) с опорным кварцевым генератором (ОКГ) и блок управления (БУ) для задания необходимых режимов работы всех модулей тракта. Синтезатор частоты (СЧ) и блок преобразования частоты (БПЧ) необходимы, если требуется работа с разными значениями промежуточной частоты либо получение радиосигнала на несущей частоте.

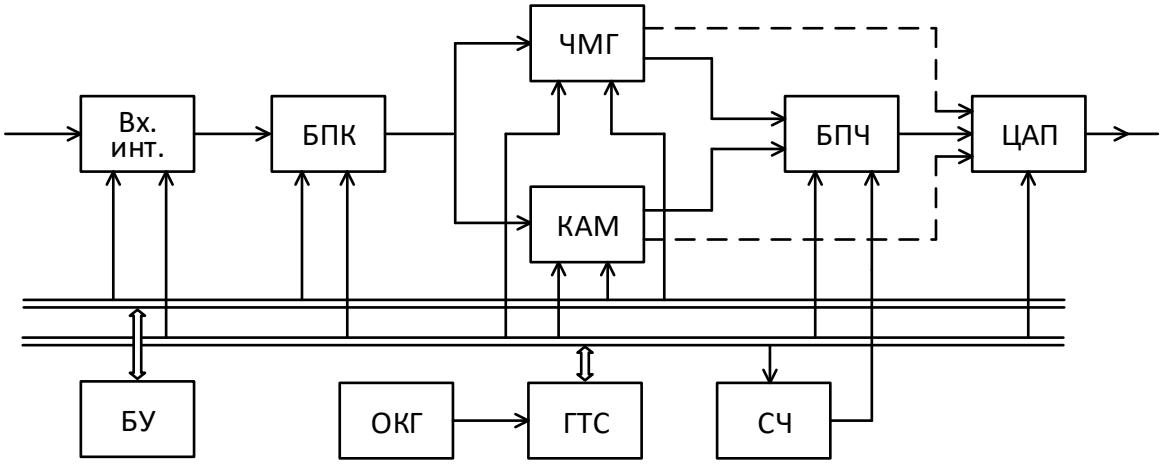


Рис. 1. Общая структурная схема универсального цифрового радиопередающего тракта

При построении многостандартных трактов одной из проблем являются разные принципы реализации модуляторов, требующие их параллельного размещения, как показано на рис. 1. Это увеличивает требования к ресурсам ПЛИС. В частности, КАМ-модуляторы реализуются на нулевой ПЧ по квадратурной схеме, а многопозиционные частотные манипуляторы строятся с использованием цифровых управляемых генераторов на базе прямого цифрового синтеза (ПЦС) частот. При этом в ряде случаев отпадает необходимость в блоках СЧ и ПЧ, так как схема ПЦС позволяет сформировать ЧМн-радиосигнал на любой частоте, не превышающей половину тактовой.

Исходя из вышеперечисленных факторов, для унификации ЧМн- и КАМ-трактов используется формирование ЧМн-сигнала на нулевой частоте. Если перенос радиосигнала на несущую частоту предполагается осуществлять аналоговым способом, то квадратурные составляющие как ЧМн-, так и КАМ-сигналов могут быть непосредственно поданы на ЦАП, как показано на рис. 1 штриховыми линиями. Использование БПЧ и СЧ в данном случае обусловлено особенностью блока ЦАП платы THDB-ADA, а именно, трансформаторным выходом, не позволяющим выдавать сигналы на нулевой ПЧ.

Подсистема формирования ЧМн-сигнала

Подсистема обеспечивает формирование ЧМн-сигнала с количеством ступеней от 2 до 1024 при шаге девиации частоты от 1 до 1024 кГц. С учетом аппаратных ограничений возможны не все сочетания, а только те, при которых максимальная девиация не превышает 8192 кГц. Символьная скорость при этом может быть установлена равной 64, 128, 256 или 512 ксимв/с. Битовая скорость при этом будет находиться в пределах от 64 до 5120 кбит/с.

Ядром подсистемы является цифровой частотно-модулированный генератор, выполненный на базе схемы прямого цифрового синтеза (ЧМГ ПЦС), выполненной по традиционной схеме с использованием фазового регистра и таблиц отсчетных значений синусов или косинусов. Отличием от схемы из [3] является использование и таблиц синусов, и таблиц косинусов, поскольку должна быть обеспечена возможность формирования ЧМн-сигнала на нулевой ПЧ. Также использованная схема гораздо проще в реализации, чем варианты с непосредственным вычислением значений тригонометрических функций, например, предлагаемые и рассматриваемые в [4], а требование наличия достаточно большого объема памяти для хранения предварительно вычисленных отсчетных значений синуса/косинуса для семейства ПЛИС Cyclone V достаточно легко выполнимо.

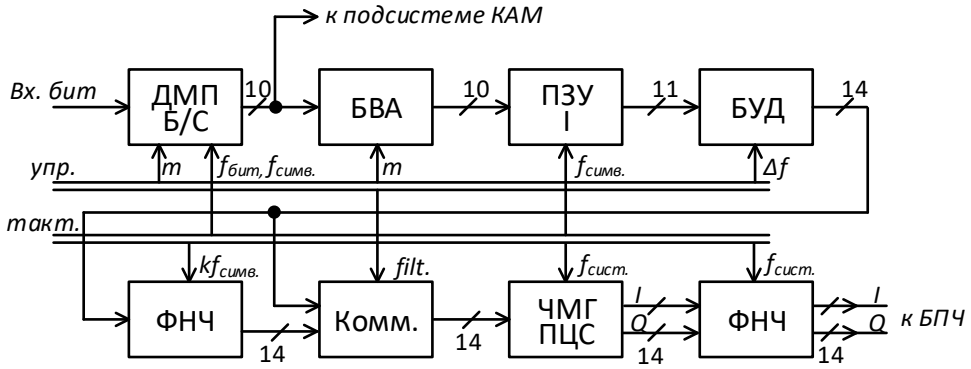


Рис. 2. Подсистема формирования ЧМн-сигнала

Значение частоты выходного синусоидального сигнала ПЦС-синтезатора определяется выражением $f_0 = \frac{Mf_s}{2^n}$, где f_s – тактовая частота работы схемы ПЦС (частота дискретизации), n – разрядность фазового регистра, M – код частоты, загружаемый в регистр, причем $0 < M < 2n - 1$.

Разрешающая способность системы по частоте равна $f_s / 2^n$. При выборе значения тактовой частоты учитывались особенности формирования сетки частот и ограничения используемой элементной базы. В первую очередь, f_s определяет диапазон частот генерируемых сигналов, которые могут лежать в пределах от $f_{MIN} = f_s / 2^n$ до $f_{MAX} < f_s / 2$. С учетом того, что амплитудно-частотная характеристика ЦАП имеет вид $K(f) = \sin(f) / f$, работа вблизи $f_s / 2$ требует применения корректирующего фильтра, а также значительно повышает требования к аналоговому восстанавливающему ФНЧ. Поскольку в плате THDB-ADA восстанавливающим фильтром является простая RC-цепь, желательным является выполнение условия $f_{MAX} << f_s / 2$.

Также от соотношения f_s и n зависит шаг сетки частот. С этой точки зрения при построении ЧМГ f_s удобно выбирать равной $\Delta f_{MIN} \cdot 2^k$, где Δf_{MIN} – минимальное значение девиации частоты. Практически же размер фазового регистра выбирается с запасом, значения 32…48 позволяют формировать сетку частот с точностью, достаточной для абсолютно всех связных приложений. В данном случае достаточен регистр длиной 32, при этом при выборе тактовой частоты $f_s = 2^{17} \cdot \Delta f_{MIN} = 131072$ кГц, являющейся максимальной для ЦАП платы THDB-ADA,
и минимальной девиации в 1 кГц, шаг сетки частот будет равным 0,031 Гц. Для выбора значения из таблиц синусов/косинусов оставлены только первые 14 старших значащих разрядов (MSB), что достаточно для обеспечения уровня внеполосных составляющих схемы не более –60 дБс. Это уменьшает размер таблиц, для хранения которых достаточно двух блоков памяти объемом $(2^{14}/4)\times 14 = 4096\times 14$ бит каждый, и не ухудшает разрешающую способность по частоте. Рассчитанный объем блоков памяти приведен с учетом вычисления отсчетных значений 2-го, 3-го и 4-го квадрантов по хранящимся в памяти значениям только 1-го квадранта, путем их инверсии или изменения направления считывания, что позволило сократить его в 4 раза. Выбор разрядности в 14 бит соответствует разрядности ЦАП и позволяет получить отношение сигнал/шум квантования не менее 70 дБ.

Для формирования сигнала управления частотой также использовано ПЗУ с предварительно вычисленными значениями. Для эквидистантной (с одинаковым шагом) ЧМн используются наборы значений от [–1:1], что соответствует 2-ЧМн, до [–1023:1023] для 1024-ЧМн. Поскольку значения для созвездий большей размерности покрывают значения меньших, для хранения отсчетов используется одно ПЗУ размером 1024×14, занимающее 2 блока встроенной памяти M10K (M9K для Cyclone IV). На рис. 3 показано расположение отсчетов в памяти для простейшего случая, без использования кода Грэя. Использование ПЗУ с предварительно вычисленными значениями существенно упрощает схему управления

частотой, при этом количество ресурсов ПЛИС не зависит от сложности вычисления кода частоты.

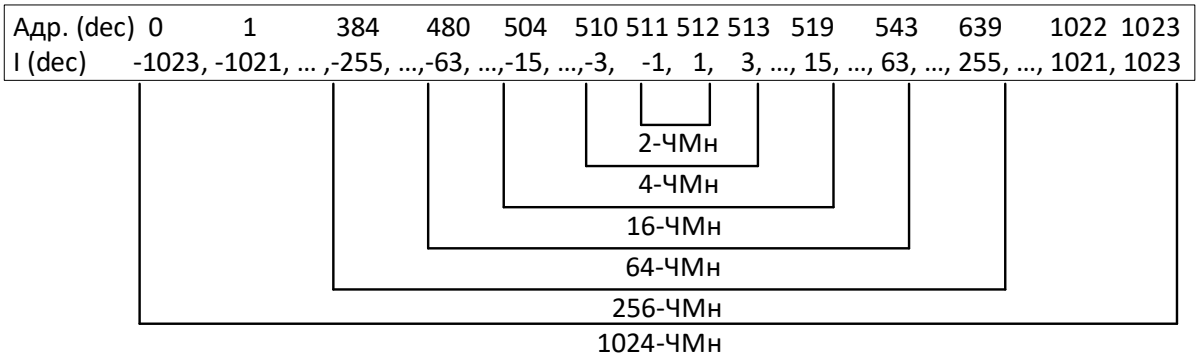


Рис. 3. Схема размещения значений сигнала управления частотой в ПЗУ

Таким образом, формирование кода управления частотой ЧМГ ПЦС осуществляется в 3 этапа: вначале в демультиплексоре (ДМП) происходит преобразование последовательного битового потока в параллельный символьный, с количеством разрядов (бит/символ) от 1 (для 2-ЧМн) до 10 (1024-ЧМн), далее в блоке вычисления адреса определяется адресное пространство для соответствующей размерности ЧМн, затем из ПЗУ считываются отсчетные значения управляющего кода, соответствующие минимальной сетке значений девиаций частот, и наконец, в блоке установки девиации частоты производится окончательное формирование управляющего кода путем битового сдвига на требуемое количество (от 0 до 9) разрядов.

Еще одной проблемой использования ЧМн-сигналов (как и любых других манипулированных сигналов) является теоретически бесконечная ширина их спектра. Даже при использовании режима манипуляции с непрерывной фазой, что и обеспечивает схема ПЦС, уровень боковых лепестков с высокой долей вероятности окажется выше допустимых значений, особенно для радиопередающих устройств с высокой выходной мощностью. Фильтр нижних частот (ФНЧ), устанавливаемый после ЧМГ, должен быть перестраиваемым (переключаемым), поскольку ширина спектра ЧМн-сигнала зависит как от символьной скорости, так и от девиации частоты. При большом количестве режимов работы ЧМГ конструкция такого фильтра может оказаться весьма сложной и ресурсоемкой. Поэтому использован дополнительный ФНЧ перед ЧМГ, ограничивающий полосу частот модулирующего сигнала. Он работает на фиксированной частоте $kf_{симв.}$, где $k = 8$ или 16 – коэффициент интерполяции. Совместное действие двух ФНЧ позволит достичь заданного уровня внеполосных излучений ЧМГ, для их реализации потребуется не более 50 умножителей, выполненных в виде выделенных DSP-модулей ПЛИС, или синтезированных из логических блоков. Последний вариант следует использовать только в случае нехватки необходимого количества выделенных модулей умножения-суммирования в малоразмерных ПЛИС.

Подсистема формирования КАМ-сигнала

Данная подсистема обеспечивает формирование сигнальных созвездий размерностью от 2 (2-ФМ) до 1024 (1024-КАМ), при этом диапазон битовых и символьных скоростей такой же, как и в ЧМн-подсистеме.

Подсистема состоит из двух идентичных каналов (рис. 4), обеспечивающих формирование комплексной огибающей КАМ-сигнала, представленной квадратурными составляющими *I* и *Q*. Последовательность символов принимается от ДМП ЧМн-подсистемы (см. рис. 2), являющегося общим для обеих подсистем. Однако в КАМ-подсистеме нечетные разряды подаются в синфазный (*I*) канал, четные – в квадратурный (*Q*), или наоборот – правило определяется для конкретного применения.

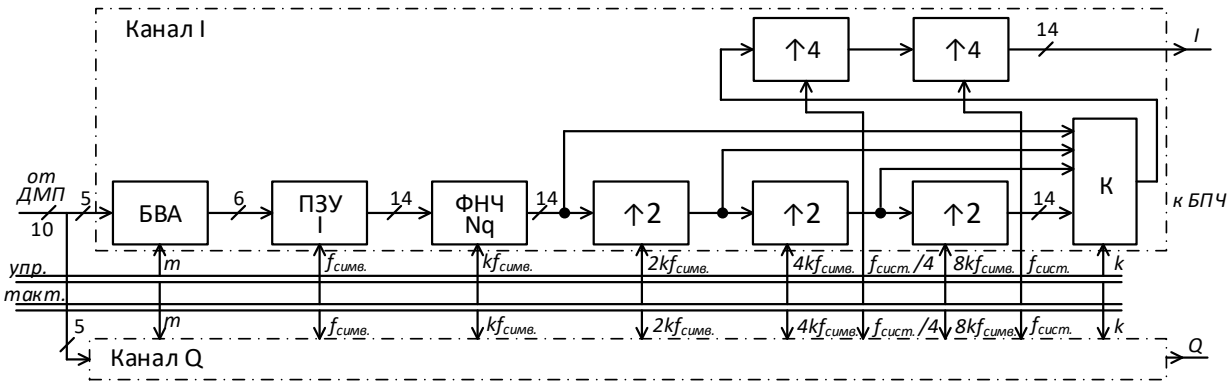


Рис. 4. Подсистема формирования ЧМн-сигнала

Для формирования *I* и *Q* сигналов использован тот же метод, что и в ЧМн-подсистеме – чтение из ПЗУ предварительно вычисленных отсчетных значений. Такое решение обусловлено необходимостью нормирования амплитуды *I* и *Q* сигналов, чтобы мощность радиосигнала на выходе модулятора была одинаковой для всех режимов. Для эквидистантных сигнальных созвездий значения сигналов будут в пределах от [–1, 1] до [–31, 31] (табл. 1).

Таблица 1. **Коэффициенты нормировки *I* и *Q* сигналов**

Вид модуляции	2-ФМн	4-КАМ	16-КАМ	64-КАМ	256-КАМ	1024-КАМ
Диапазон значений сигналов *I* и *Q*	–1, 1	–1, 1	–3, –1, 1, 3	–7, …, –1, 1, …, 7	–15, …, –1, 1, …, 15	–31, …, –1, 1, …, 31
Коэффициент нормировки K_n	1 (0 для *Q*)	$1/\sqrt{2}$	$1/\sqrt{10}$	$1/\sqrt{42}$	$1/\sqrt{170}$	$1/\sqrt{682}$
Множитель K_x	4095 (0)	2896	1295	632	314	157

Коэффициенты нормировки являются, в основном, иррациональными числами, поэтому переводятся в целочисленные с учетом выбранной разрядности *n* формируемых *I* и *Q* сигналов, в настоящем случае 14 бит. Множитель (коэффициент пересчета) вычисляется по выражению $K_x = round((2^{n-1}-1)/K_n)$.

Поскольку диапазон значений сигналов *I* и *Q* и множители заранее известны, то они могут быть вычислены предварительно и записаны в ПЗУ (непосредственно, в коде Грэя или в соответствии с любым иным правилом формирования сигнального созвездия), в результате чего схема упрощается – из нее исключается кодер Грэя и умножители на коэффициент. Значения *I* и *Q* сигналов для различных видов модуляции записываются в одно ПЗУ, для чего его адресное пространство разделяется в соответствии с табл. 2.

Таблица 2. **Разделение адресного пространства ПЗУ хранения отсчетов *I* и *Q* сигналов**

Вид модуляции	2-ФМн	4-КАМ	16-КАМ	64-КАМ	256-КАМ	1024-КАМ
Диапазон адресов	0	1…2	4…7	8…15	16…31	32…63

Для реализации разделения адресного пространства в блоке вычисления адреса (БВА) к коду сигнала добавляется значение от 2^0 (для 4-КАМ) до 2^5 (для 1024-КАМ). Емкость ПЗУ составляет 64×14 бит, содержание для *I* и *Q* каналов идентично за исключением первой ячейки – в ПЗУ *I* канала оно равно 4095, в ПЗУ *Q* канала – 0 (см. табл. 1).

Ограничение полосы частот выполняет ФНЧ с кососимметричным склоном АЧХ – фильтр Найквиста [2]. В зависимости от требований к крутизне склона, неравномерности АЧХ в полосе пропускания, затухания в полосе задержания и соотношения частоты среза к частоте дискретизации КИХ-фильтр Найквиста будет иметь порядок от 20 и более, что является достаточно ресурсоемким с точки зрения использования DSP-блоков ПЛИС. Также как и в ЧМн-подсистеме, фильтр работает на частоте $kf_{симв.}$, *k* принято равным 16. Таким

образом, одновременно выполняется 16-кратная интерполяция методом повторения соседних отсчетов.

Поскольку с выхода фильтра данные поступают с тактовой частотой, зависящей от выбранной символьной скорости и принимающей значения от 1024 кГц (при 64 ксимв/с) до 8192 кГц (при 512 ксимв/с), для ее выравнивания интерполятор выполнен по каскадной схеме – вначале трех каскадный коммутируемый, каждая ступень которого обеспечивает двукратную интерполяцию, затем двух каскадный фиксированный, с четырех кратной интерполяцией. В результате выходная тактовая частота составляет 131072 кГц независимо от символьной скорости. Все интерполяторы являются каскадными интегрирующе-гребенчатыми (CIC) фильтрами, не требующими при реализации операций умножения. Разделение интерполятора на каскады позволяет последовательно снижать требования к каждой секции, соответственно, оптимальнее использовать ресурсы ПЛИС.

Подсистемы синтеза и преобразования частот, тактирования и управления

Синтезатор частоты для БПЧ (см. рис. 2) также выполнен на базе схемы ПЦС, его параметры полностью идентичны параметрам синтезатора, использованного в ЧМн-подсистеме. Соответственно, значение ПЧ может быть от 0 до 65 МГц, однако практически, с учетом вышерассмотренных особенностей платы ЦАП, его не следует выбирать выше 20…30 МГц.

Преобразователь частоты – классический квадратурный. Его цифровая реализация позволяет добиться математически точного формирования радиосигнала, проблемы разбаланса фаз и амплитуд квадратурных каналов, присущие аналоговым схемам, за счет выбора большой разрядности вычислений (14 и более бит), могут быть сведены до пренебрежимо малых.

Для формирования необходимых тактовых сигналов используется встроенная в ПЛИС схема ФАПЧ, но поскольку значения частот сигналов символьной синхронизации лежат ниже ее минимально возможных (т. е. ниже 1 МГц), то используются дополнительные делители частоты.

Блок управления выполняет декодирование сигналов от кнопочной панели управления и формирование на их базе необходимых управляющих сигналов для подсистем модулятора.

Заключение

Разработанный универсальный модулятор позволяет формировать радиосигналы практически со всеми используемыми видами ЧМн и КАМ, реализуем на любой ПЛИС семейства Cyclone V, также может быть легко перенесен на другие ПЛИС при условии наличия у них достаточного количества встроенных модулей памяти и DSP-блоков. Основное практическое применение для него – отработка алгоритмов формирования радиосигналов для цифровых систем связи. При использовании модулятора в системах, не требующих столь широкой номенклатуры формируемых сигналов, конструкция может быть упрощена.

UNIVERSAL FSK/QAM MODULATOR ON ALTERA CYCLONE V FPGA

A.L. KHAMINICH

Abstract. Implementation on a programmable logic integrated circuit (FPGA) of the Intel/Altera Cyclone V family of a digital modulator, which provides the formation of frequency- and quadrature modulated radio signals with a signal constellation dimension from 2 to 1024 us considered. It's suggested options for the modulators implementation that ensure efficient use of FPGA resources.

Keywords: keying, modulation, interpolation, signal constellation, frequency conversion.

Список литературы

1. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. М.: Издат. дом «Вильямс», 2003.
2. Прокис Дж. Цифровая связь. М.: «Радио и связь», 2000.
3. Аналого-цифровое преобразование / под ред. У. Кестера. М.: Техносфера, 2007.
4. Ильинков В.А., Ярков Я.М., Ильинкова А.В. // Докл. БГУИР. 2016. № 5 (99). С. 24–29.

UDC 004.021

NEW ROBUST FACE ANTI-SPOOFING TECHNIQUE

Z.T. KHUDOYKULOV, Sh.Z. ISLOMOV, L.U. DAVRONOVA, U.R. MARDIEV

Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Republic of Uzbekistan

Submitted 20 March 2019

Abstract. Face recognition is one of the main areas of biometric identification and authentication. Face identification consists of face detection and recognition processes. There are a lot of attacks which they can spoof of faces on face detection process. In this paper all attack points between camera and user application are given. According to the attack points, there are three types of anti-spoofing methods. Based on feature level techniques robust anti-spoofing technique is proposed. When intruder uses photo of the registered person, proposed anti-spoofing technique easily detects. This anti-spoofing technique detects fake face based on pixel value between face and background space.

Keywords: face detection, face recognition, face identification, spoofing, sensor, anti-spoofing, score-level.

Introduction

Identification and authentication techniques based on face are widely used in access control systems of organization. Face identification consists of two processes: Face detection, Face recognition. Face detection is a process finding face gradients and extraction faces from images. Based on detected faces are recognized person. Face detection and recognition processes is presented in Fig. 1.

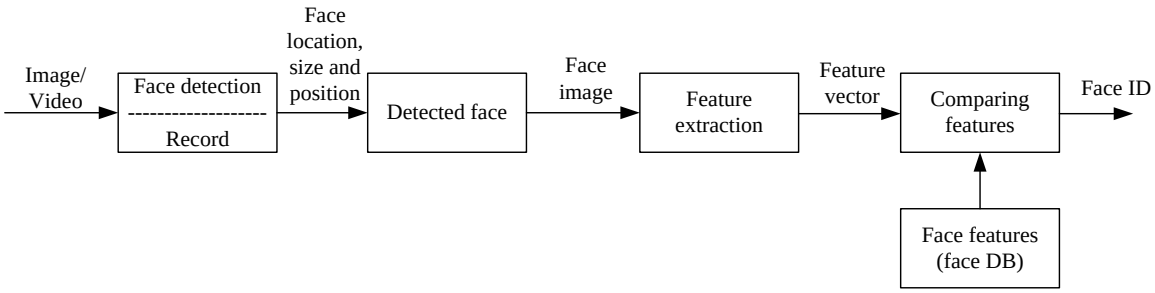


Fig. 1. Face detection and recognition processes

Here, first image is captures from web camera or video. Face detection detects faces from image and features which are extracted from detected faces. Extracted features compares with stored parameters in face database (face DB).

Attack points

These processes are presented without any attacks. Also, there are a lot of types of attacks which is disturbed identification systems. In [1] eight attack points between scanner and identification applications is given. The main components of face identification system are given in Fig. 2.

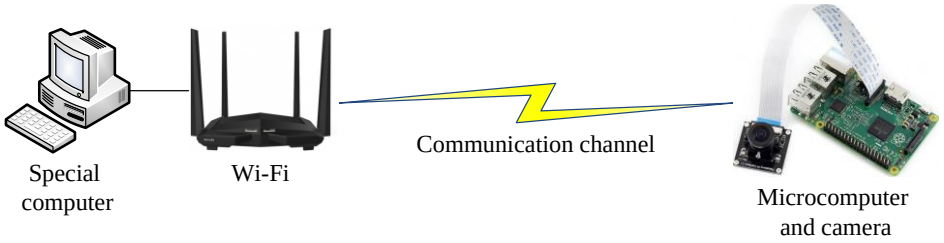


Fig. 2. Face identification system using microcomputers

The camera captures image and Wi-Fi, which supports microcomputer send it over the wireless communication channel. A special computer, which is used to detect faces and extract facial features to identify a person. Here, attack points are in web camera, communication channel and special computer. In Fig. 3 all attack points are given.

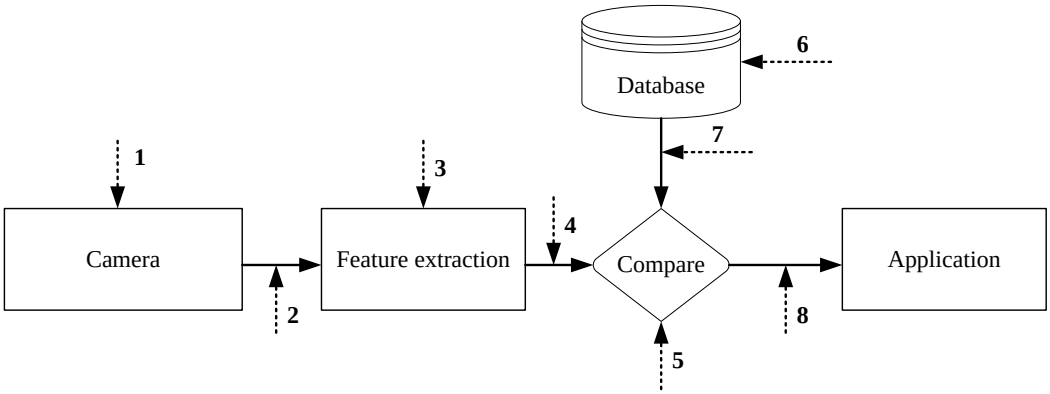


Fig. 3. Attack points

Point 1 attacker sends fake face image (photo, video or 3D mask) to the camera. Camera takes fake face image and sends to the special computer over the communication channel. For protection from spoofing faces anti-spoofing methods are used [2].

Point 2 captured image from camera sends to the special computer for preprocessing over the wireless channel. Channel is protected with cryptographic algorithms, but there are many cryptanalysis methods [3], which can interrupt encrypted packets. For protection from these attacks types strong and combined cryptographic methods are used.

Point 3, 4, 5, 6, 7 is directed to attack computer systems, programs and databases. Monitoring, filtering and access control tools are used for protection.

Point 8 these types of direct attacks to change packets, which saved identification results (like point 2).

Proposed method

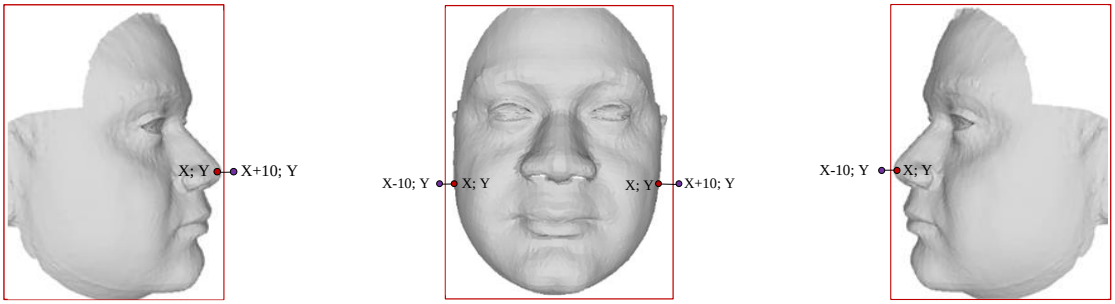
Three types of detection methods from attacks are used to detect spoofed faces: sensor, feature and score level techniques.

In sensor-level techniques special hardware to detect spoofing is used. These methods add some special device to the sensor in order to detect particular properties of a living trait (e.g., facial thermogram, blood pressure, fingerprint sweat, or special reflection properties of the eye).

In feature-level techniques special software to distinguish between real and fake traits which are extracted from the biometric sample are used. For instance, from just one single high resolution image of a face we may perform both face and iris recognition. In this particular case, a multimodal strategy is being applied at the feature extractor level, with no need of any additional hardware or sensing device.

Score-level techniques are used when two-type detection methods fall out. This technique is used in score level. For example, second type of biometric method is used for detection spoofing, or studied results of other type of anti-spoofing methods.

Proposed anti-spoofing pixel-based method is developed on feature-level techniques. When an attacker tries to use fake face photo, pixel based anti-spoofing method detects it easily. In this method comparing the background of the detected face and face gradients is used. In real face identification systems without spoofing background image and face gradients changes over time. If these pixels do not change, it means that face image is spoofed.



a *b* *c*

Fig. 4. Different views of faces: right (*a*), front (*b*), left (*c*)

In [4] triangle method for faster and accurately face detection is proposed. After detection face regions, we can distinguish pixel values between face space and background space. The value of *X* coordinate from face and value of *X*+10 coordinate from background comparison is shown. 3 positions of the faces are viewed in this method.

1. *Right.* In left side-face, in right side-background image pixels are situated (Fig. 4, *a*). Here, $A(X;Y)$ is detected from face space and $B(X+10;Y)$ is detected from background space. By $I=|A-B|$ formula is calculated differences between *A* and *B*. After 0,5 second is computed $I_1=|A-B|$. If following condition is done, then this face is real, else, spoofed face.

$|I - I_1| \geq 10$;

$A(X,Y)$ point in face;

$B(X+10,Y)$ point in background image;

$I=|A-B|$;

$I_1=|A-B|$ after 0,5 second;

$|I - I_1| \geq 10$ is real face, else fake.

2. *Front*. Here is applied Right and Left methods (Fig. 4, *b*).

3. *Left*. In right side-face, in left side-background image pixels are situated (Fig. 4. *c*). Here, $A(X,Y)$ is detected from face space and $B(X-10,Y)$ is detected from background space. By $I=|A-B|$ formula calculates differences between *A* and *B*. After 0,5 second is computed $I_1=|A-B|$. If following condition is done, then this face is real, else, spoofed face.

Comparison and discussions

Proposed anti-spoofing method was analyzed with IQA [5], IDA [6] and DNN [7] methods. IQA (Image Quality Assessment) detects and finds spoofed faces based on image quality. IDA (Image Distortion Analysis) detects spoofed faces based on image distortion. DNN (Deep Neuron Network) detects spoofed faces based on spoof face databases. DNN presents high result, but it takes more time and resource to train face databases, also, it is difficult to collect spoofed faces. Presentation of accuracy of anti-spoofing techniques is in table 1.

Table 1. **Comparison of anti-spoofing techniques**

Method	IQA	IDA	DNN	Proposed method
Accuracy	0,86	0,9	0,98	0,95

Conclusion

In one of the situations, the spoofed face was detected, and it was a fake face image. By applying this method for detecting fake faces in face recognition, the system is protected from spoofing without any special hardware or second type of biometrics.

References

1. Rubal J., Chander K. // Int. J. of Advances in Scientific Research 2015. Vol. 1 (07) P. 283–288.
2. Galbally J., Marcel S., Fierrez J. // IEEE Access. 2014. T. 2. C. 1530–1552.
3. Moabalobelo T., Nelwamondo F., Tsague H. D. Survey on the cryptanalysis of wireless sensor networks using side-channel analysis. 2012.
4. Malikovich K.M. [et al.] // The J. of Multimedia Information System. 2018. Vol. 5. №. 1. P. 15–20.
5. Hanimol T.M., Vidhushavarshini S. Bio-Spoof and Anti-Spoofing Detection Based on IQA. 2015.
6. Wen D., Han H., Jain A.K. // IEEE Transactions on Information Forensics and Security. 2015. Vol. 10. №. 4. P. 746–761.
7. Wang M., Deng W. Deep face recognition: a survey. 2018.

СЕКЦИЯ 2
ОБРАБОТКА И ПЕРЕДАЧА ИЗОБРАЖЕНИЙ

УДК 681.3.06

КОДИРОВАНИЕ И СЖАТИЕ СЕГМЕНТИРОВАННОГО ИЗОБРАЖЕНИЯ

А.И. МИТЮХИН

Институт информационных технологий Белорусского государственного университета информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 марта 2019

Аннотация. В статье рассматривается алгоритм эффективного представления изображения, полученного после процесса сегментации. Предлагается комбинированный метод описания сегментированного объекта, включающий в себя выполнение двух последовательных этапов цифровой обработки исходных данных. Приведена сравнительная оценка эффективности сжатия и информационного содержания исходных данных с применением цепного кода Фримена, энтропийного кодирования и рассматриваемого комбинированного алгоритма. Представлен пример сокращения размерности обработки изображения сегментированного объекта.

Ключевые слова: изображение, граница объекта, сегментация, энтропия, форма, последовательность, окрестность, источник.

Введение

Операции распознавания некоторой изолированной области, присутствующей на изображении (снимке), предшествует этап сегментации. Желательно, чтобы число признаков, описывающих процесс сегментации, было минимально возможным. При этом, для повышения вероятности правильного последующего процесса распознавания, описание сегментированного объекта должно выполнятся с определенной степенью точности. В статье рассматривается математический подход, основанный на комбинировании уже существующих алгоритмов обработки изображений для достижения уменьшения вычислительных затрат на представление и хранение данных, являющих объектом интереса

Теоретические принципы

Пусть имеется снимок $f(x,y)$ отдельного объекта определенной формы. В результате сегментации на дискретной сетке в пространственной области получено бинарное изображение $g(x,y)$. Изображение можно записать в виде 2D-последовательности (вектора):

$$g = g(x,y) = ((x_0,y_0),...,(x_n,y_n)), \quad (1)$$

где (x,y) – пространственные целочисленные переменные, n – длина границы, $n \in Z^+$. Пикселы с координатами (x_i,y_i) и (x_{i-1},y_{i-1}) удовлетворяют свойству m-компонентной окрестности, $0 \le i \le m$. При значительной величине n последовательности g, а также при необходимости выполнения требования точной сегментации и высокого разрешения вычислительные затраты на представление g могут оказаться недопустимыми.

Для определенных приложений, например, хранения отдельных деталей изображаемых объектов интереса, 2D-последовательность g можно представить в виде случайной последовательности $g(x,y)$ символов дискретного источника информации без памяти:

$$G=\{g_0,g_1,...,g_m\}, \quad (2)$$

где m определяет вид окрестности (на практике используется 4-компонентная или 8-компонентная окрестность). Источник, формирующий g, полностью описывается множеством $P=\{p(g_0),p(g_1),...,p(g_m)\}$ вероятностей символов. Знание распределения множества вероятностей P позволяет рассчитать энтропию источника G.

$$H=-\sum_{i=0}^{m}p(g_i)\log p(g_i). \quad (3)$$

В свою очередь, значение H определяет верхнюю и нижнюю границы средней длины l эффективного q-ичного префиксного кода, с помощью которого можно выполнить сжатие данных. В результате появляется возможность получения точной оценки энтропийного кодирования источника, эффективности сжатия. Недостатком энтропийного алгоритма кодирования является сравнительно низкий коэффициент сжатия:

$$L=\frac{N}{M}=\frac{N}{nl}\le 3, \quad (4)$$

где N – затраты в виде числа носителей информации (например, двоичных чипов) на представление, хранение исходных данных без сжатия, M – затраты в виде количества чипов необходимых на представление, хранение исходных данных со сжатием.

Известно, что определенная эффективность описания контуров, замкнутых границ, линий, достигается с использованием цепного кода Фримена [1]. Кодирование состоит в присваивании отрезку контура (границы, линии) одноразрядного десятичного числа в соответствии с направлениями, показанными на рис. 1. Следует отметить как достоинство метода применение цепного кода, это то, что для пространственной «синхронизации» процесса декодирования (восстановления изображения) достаточно знания координат (x_i,y_i) только одного пиксела. Этот пиксел задает начальную пространственную фазу кода.

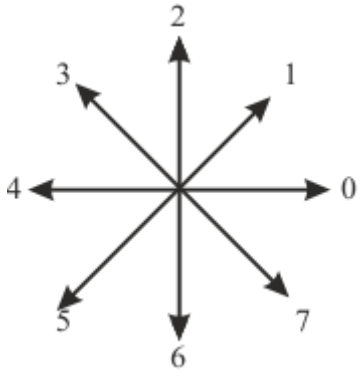


Рис. 1. Кодирование векторов направлений в 8-компонентной окрестности

Уменьшение вычислений при обработке изображений с применением цепного кода происходит за счет кодирования исходных данных равномерными словами длиной два или три двоичных символа (чипа). Таким образом, начальный этап обработки изображения включает в себя сегментацию изображения $f(x,y)$ для получения последовательности $g=g(x,y)$. Далее последовательность кодируется цепным кодом $c=c_0,c_1,...,c_n$. В статье рассматривается применимость к цепному коду линейного ортогонального преобразования, позволяющего повысить эффективность сжатия исходных данных. Известно, среди всех линейных преобразований оптимальным по отношению к критерию наименьшего значения

среднеквадратической ошибки при неполной размерности (неточном задании координат) является декоррелирующее преобразование в базисе дискретных функций [1, 2]. Рассматривается подход, основанный на преобразовании цепного кода $c = c_0, c_1, ..., c_n$ в систему координат, базовым векторным пространством которого служат собственные векторы ковариационной матрицы cov($\boldsymbol{c}$) цепного кода $\boldsymbol{c}$. Вычисление вектора собственных значений ковариационной матрицы cov($\boldsymbol{c}$) позволяет произвести оценку возможной эффективности сжатия обрабатываемых данных:

$$\boldsymbol{\Lambda} = (\lambda_0, \lambda_1, ..., \lambda_n) \qquad (5)$$

При векторном описании сегментированного изображения, соответственно, кода $\boldsymbol{c}$ как выхода источника (2), можно утверждать, что среднее количество информации этого источника, приходящееся на символ, равно энтропии (3). Так как значение энтропии не зависит от того, в каком порядке следуют символы источника, кодовую последовательность c запишем в виде матрицы $\mathbf{C}$ размером $(n \times 2)$. Из практических и вычислительных соображений временные затраты на обработку входного процесса всегда уменьшаются, если заранее исключать составляющею входа отражающую математическое ожидание. Тогда преобразованию будут подлежать входные данные:

$$\mathbf{B} = \mathbf{C} - E(\mathbf{C}), \qquad (6)$$

где $E(\mathbf{C})$ – среднее значение матрицы $\mathbf{C}$.

Обозначим матрицу собственных векторов через $\mathbf{A}$. Прямое преобразование матрицы $\mathbf{B}$ можно записать следующим образом:

$$\hat{\mathbf{B}} = \mathbf{A}^T\mathbf{B}, \qquad (7)$$

где $\mathbf{A}^T$ – ядро преобразования в базисе собственных векторов.

Имея распределение значений дисперсий (5) коэффициентов преобразования (7), можно решить задачу уменьшения размера входа обработки. Матрицы $\mathbf{A}$ и $\mathbf{B}$ представляются меньшим числом коэффициентов. К этим укороченным матрицам применяется обратное преобразование:

$$\mathbf{B}'' = \mathbf{A}''\hat{\mathbf{B}}'', \qquad (8)$$

где $\mathbf{A}''$ – укороченная матрица ядра обратного преобразования $\mathbf{A}$, $\hat{\mathbf{B}}''$ – укороченная матрица, полученная из матрицы коэффициентов преобразования $\hat{\mathbf{B}}$. В результате достигается уменьшение размера входа обработки данных.

Пример. Кодирование изображения цепным кодом и сжатие.

На рис. 2. показано изображение кафедрального собора. Требуется компактно представить форму собора, используя его граничные пиксели. После выполнения сегментации объекта и кодирования с помощью 8-компонентной окрестности получен двумерный цепной код c длиной $n = 42$. Элементы кода показаны в таблице. Кодовой последовательности c соответствует матрица $\mathbf{C}$ размером (42×2).

Цепной код сегментированного объекта

№ пиксела	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Элементы цепного кода	0 6	7 6	6 7	7 6	6 6	6 7	6 6	6 6	6 6	6 6	6 6	7 0	0 7	7 7
№ пиксела	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Элементы цепного кода	7 6	6 6	6 6	4 4	4 4	4 4	4 4	3 4	4 4	4 4	4 5	4 4	4 2	2 2
№ пиксела	28	29	30	31	32	33	34	35	36	37	38	39	40	41
Элементы цепного кода	2 2	2 1	1 1	1 2	2 2	2 2	2 2	2 2	2 2	1 2	2 2	2 1	2 1	2 2

Рис. 2. Изображение кафедрального собора

Вычислив среднее значение $E(\mathbf{C})$ и используя выражение (6), находим матрицу входа с нулевым математическим ожиданием **B**. Далее вычисляется ковариационная матрица cov(**B**) размером 42×42. По матрице cov(**B**) вычисляется распределение $\mathbf{\Lambda}$ собственный значений λ_i

$$\mathbf{\Lambda} = (\lambda_0, \lambda_1, ..., \lambda_{41}) = (0, 0, ..., 75).$$

Как видно, имеется только одно ненулевое собственное значение равное 75. Таким образом, возможна полная декорреляция исходных данных и тем самым компактное их представление и описание для определенных практических применений. Вычисление матрицы собственных векторов **A** ковариационной матрицы cov(**B**) позволяет определить по выражению (7) матрицу коэффициентов преобразования:

$$\hat{\mathbf{B}} = \begin{pmatrix} \hat{b}_{00} & \hat{b}_{10} & ... & \hat{b}_{410} \\ \hat{b}_{10} & \hat{b}_{11} & ... & \hat{b}_{141} \end{pmatrix}^T = \begin{pmatrix} 0 & 0 & ... & 6{,}1237 \\ 0 & 0 & ... & -6{,}1237 \end{pmatrix}^T. \qquad (9)$$

Из (9) следует, что для описания данных в области преобразований, достаточно сохранить значения коэффициентов $\hat{b}_{4,10}$ и $\hat{b}_{1,41}$. В этом случае затраты на хранение изображения (без информации о координатах начального пиксела) составят величину $M = 2$. Промежуточное обратное преобразование вычисляется по формуле (8). Точное восстановление сегментированной границы реализуется с использованием вектора:

$$\hat{\mathbf{b}}'' = (6{,}1237 - 6{,}1237)^T, \qquad (10)$$

среднего вектора $E(\mathbf{C})$ в области исходных данных (всегда можно вычислить заранее) и только одной строки $\mathbf{A}''$ матрицы ядра обратного преобразования **A**.

Проведем сравнительную оценку информационного содержания исходных данных и эффективность сжатия с применением цепного кода Фримена, энтропийного кодирования и рассматриваемого комбинированного алгоритма.

1. Кодирование граничных пикселов цепным равномерным кодом на основе 8-компонентной окрестности потребует $N = 2nl = 252$ двоичных символов (чипов).

2. Вычисление энтропии дискретного изображения (см. таблицу и (3)), когда учитываются вероятностные характеристики источника (2)

$P = \{p(g_0),..., p(g_7)\} = \left\{\frac{1}{42}, \frac{7}{84}, \frac{2}{7}, \frac{1}{84}, \frac{17}{84}, \frac{1}{84}, \frac{11}{42}, \frac{7}{84}\right\}$, дает следующее значение:

$H = 2{,}1789 \frac{\text{бит}}{\text{символ}}$.

Полученное значение соответствует предельному минимальному значению средней длины $\bar{l}$ кода. На практике средняя длина $\bar{l}$ слова энтропийного кодирования находится в пределах $H+1>\bar{l}>H$. В наилучшем случае это означает, что исходные данные могут быть представлены с помощью $M\approx 183$ двоичных чипов. В сравнении с равномерным кодированием цепным кодом, коэффициент сжатия (выигрыш в записи данных) $L_{entropy}$ (4) в описании составляет величину

$$L_{entropy}=\frac{N}{M}=1{,}377.$$

3. Комбинированный метод описывается с помощью 8-битового двоичного кода. На описание элементов вектора $\hat{\mathbf{b}}''$ (9) и двух координат начального пиксела границы требуется $M'=32$ чипов. Коэффициент сжатия равен 7,785.

Как видно, предлагаемый метод обеспечивает сравнительно высокую степень сжатия с нулевой ошибкой восстановления изображения после процедуры декодирования.

Заключение

Подход на основе объединения метода кодирования сегментированных изображений объектов посредством цепного кода и метода ортогонального преобразования в базисе дискретных собственных функций позволяет осуществлять эффективное описание данных. Применение комбинированного алгоритма для приложений, связанных с распознаванием, хранением значительных информационных массивов, позволяет снижать технические затраты на хранение данных, более экономно использовать вычислительный ресурс, уменьшать энергетические затраты.

ENCODING AND COMPRESSING THE IMAGE SEGMENTED

A.I. MITSIUKHIN

Abstract. An algorithm for effective representation of the image obtained after segmentation is described. The combined method for describing segmented object including two successive phases of digital processing of source data is proposed. The comparative evaluation of the effectiveness of the compression and information content of the raw data using chain code Freeman, entropy coding and reporting combined algorithm is given. An example of reducing dimension processing segmented object is shown.

Keywords: image of the object boundary, segmentation, entropy, form, consistency, a neighborhood source.

Список литературы

1. Jahne B. Digital Image Processing. Concepts. Algorithms, and Scientific Applications. E-bok, 2013.
2. Mitsiukhin A. // Proceedings 59th IWK. Ilmenau Scientific Colloquium, Technische Universität Ilmenau, Band 59, 2017, Heft 2.2.02, 6 Seiten.

УДК 621.391

ВСТРЕЧНОЕ ВОЛНОВОЕ ВЫРАЩИВАНИЕ ОБЛАСТЕЙ ЛОКАЛЬНЫХ ЭКСТРЕМУМОВ ПОЛУТОНОВЫХ ИЗОБРАЖЕНИЙ

А.Т. НГУЕН, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 11 марта 2019

Аннотация. Предложен алгоритм сегментации полутоновых изображений на основе встречного волнового выращивания областей локальных экстремумов (минимумов и максимумов). В отличие от известных алгоритмов сегментации предложенный алгоритм позволяет разделять области с плавными перепадами яркости, адаптироваться к ограниченному времени сегментации и контролировать количество сегментов.

Ключевые слова: локальный экстремум, сегментация, волновое выращивание областей.

Введение

Сегментация изображений часто используется в обработке видеоинформации. В некоторых случаях время сегментации может быть ограничено, либо необходимо контролировать количество сегментов изображения. Сегментация на основе квантования по гистограмме не обеспечивает точное разделение областей из-за присвоения одинаковых номеров сегментам с одинаковой яркостью. Кроме того, известные методы сегментации, основанные на формировании областей с использованием водораздела [1–4], квантовании по гистограмме [5], разделении и слиянии областей с использованием квадра-дерева [6–8], выращивании областей [9–12], не эффективны для разделения областей с плавными перепадами яркости. Все рассмотренные методы не обеспечивают адаптацию к ограничениям на время сегментации и не позволяют контролировать количество сегментов. В этой связи актуальна задача разработки метода сегментации изображений, учитывающего перечисленные недостатки.

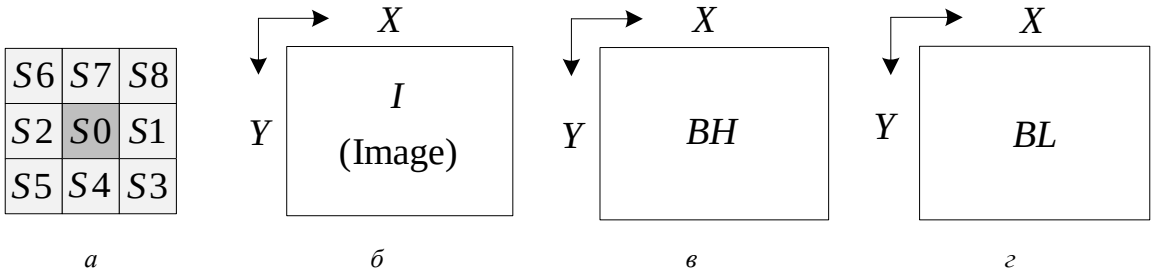
Целью работы являлось создание метода сегментации изображений, позволяющего разделять области с плавными перепадами яркости, адаптироваться к ограничениям на время сегментации и контролировать количество сегментов.

Описание алгоритма волнового выращивания областей на основе поиска экстремумов

Предлагается алгоритм сегментации полутоновых изображений на основе встречного волнового выращивания областей локальных экстремумов (минимумов и максимумов). Сущность алгоритма состоит в выделении на изображении локальных экстремумов, постепенном присоединении к ним новых элементов с учетом их местоположения вокруг экстремумов и значения порога, изменяемого от значения экстремума в противоположном направлении. Процесс волнового выращивания продолжается до тех пор, пока все области не будут выделены. Отличие предложенного алгоритма от известного алгоритма выращивания областей заключается во встречном квазипараллельном увеличении размеров выделенных областей вокруг максимумов и минимумов, что позволяет повысить точность сегментации изображений с плавными перепадами яркости. Предложенный алгоритм отличается от алгоритма волнового выращивания областей [13, 14] и регрессивного волнового выращивания областей [15, 16] выбором начальных точек роста областей в результате поиска локальных экстремумов изображения, и увеличением областей за счет постепенного присоединения к ним соседних элементов с учетом значений порогов, изменяемых постепенно на интервале между смежными локальными экстремумами противоположных знаков. Это позволяет повысить

точность и устойчивость сегментации, а также контролировать количество сегментов изображений с плавными перепадами яркости.

Исходными данными для алгоритма являются матрица изображения $I(Y,X)$, матрицы блокировки максимумов $BH(Y,X)$ и минимумов $BL(Y,X)$, сканирующая маска для поиска экстремумов (рис. 1).



а *б* *в* *г*

Рис. 1. Сканирующая маска и матрицы обработки: *а* – сканирующая маска; *б* – матрица изображения; *в* – матрица блокировки максимумов; *г* – матрица блокировки минимумов

Алгоритм сегментации изображений на основе встречного волнового выращивания областей локальных экстремумов состоит из следующих шагов.

1. Поиск экстремумов полутоновых изображений.

1.1. Если $BH(y,x) \leftarrow 0$ – пропуск $S0(y,x)$, пропуск поиска максимума. Если $BL(y,x) \leftarrow 0$ – пропуск $S0(y,x)$, пропуск поиска минимума.

1.2. Если $BH(y,x) \leftarrow 1$ – проверка $S0$ на максимум относительно $S1$, $S2$, $S3$, $S4$, $S5$, $S6$, $S7$, $S8$.

Если $S0 > Si$, $i = \overline{1,8}$, то $S0$ – максимум (однопиксельный), тогда $BL(y,x) \leftarrow 0$ и

$$\begin{cases} BH(S1) \leftarrow 0, \\ BH(S3) \leftarrow 0, \\ BH(S4) \leftarrow 0, \\ BH(S5) \leftarrow 0. \end{cases} \quad (1)$$

Иначе, если $(S0 = S1) \vee (S0 = S3) \vee (S0 = S4) \vee (S0 = S5)$, то осуществляется проверка области на максимум с использованием выращивания [9–12] и уточнения областей.

Иначе, $BH(y,x) \leftarrow 0$ и

$$\begin{cases} BL(S1) \leftarrow 0, \\ BL(S3) \leftarrow 0, \\ BL(S4) \leftarrow 0, \\ BL(S5) \leftarrow 0. \end{cases} \quad (2)$$

1.3. Иначе, если $BL(y,x) \leftarrow 1$ – проверка $S0$ на минимум относительно $S1$, $S2$, $S3$, $S4$, $S5$, $S6$, $S7$, $S8$.

Если $S0 < Si$, $i = \overline{1,8}$, то $S0$ – минимум (однопиксельный), тогда $BH(y,x) \leftarrow 0$ и

$$\begin{cases} BL(S1) \leftarrow 0, \\ BL(S3) \leftarrow 0, \\ BL(S4) \leftarrow 0, \\ BL(S5) \leftarrow 0. \end{cases} \quad (3)$$

Иначе, если $(S0 = S1) \vee (S0 = S3) \vee (S0 = S4) \vee (S0 = S5)$, то осуществляется проверка области на минимум с использованием выращивания [9–12] и уточнения областей.

Иначе, $BL(y,x) \leftarrow 0$ и

$$\begin{cases} BH(S1) \leftarrow 0, \\ BH(S3) \leftarrow 0, \\ BH(S4) \leftarrow 0, \\ BH(S5) \leftarrow 0. \end{cases} \quad (4)$$

В результате выполнения пункта 1 формируется матрица $EM = \left\| em(y,x) \right\|_{(y=\overline{0,Y-1},x=\overline{0,X-1})}$ экстремумов, значение каждого элемента которой указывает на номер экстремума или его отсутствие. Эти данные используются для встречного волнового выращивания областей.

2. Сегментация на основе встречного волнового выращивания областей.

2.1. Формирование матрицы $MT = \left\| mt(y,x) \right\|_{(y=\overline{0,Y-1},x=\overline{0,X-1})}$ поиска 4-связных соседей, элементы которой определяются с помощью выражения $mt(y,x) \leftarrow 0$ при $y = \overline{0,Y-1},\ x = \overline{0,X-1}$.

2.2. Инициализация стеков $SY = \left\| sy(k) \right\|_{(k=\overline{0,K-1})}$ и $SX = \left\| sx(k) \right\|_{(k=\overline{0,K-1})}$ для временного хранения координат соседей, элементы которых определяются с помощью выражения $sy(k) \leftarrow 0$ и $sx(k) \leftarrow 0$. Указатель стеков устанавливается в 0: $k \leftarrow 0$.

Указатель флага $Flag$ для остановки цикла сегментации устанавливается в единицу: $Flag \leftarrow 1$.

2.3. Начало цикла сегментации.

2.3.1. В окрестности экстремумов осуществляется поиск 4-связных соседей, координаты которых заносятся в стеки SY и SX. Присвоение соответственных номеров элементам матрицы MT описывается с помощью выражения

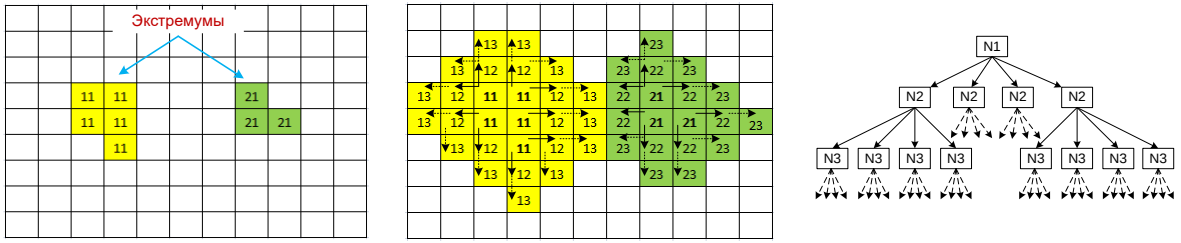
$$\begin{gathered} \exists(Flag = 1) \Rightarrow Flag \leftarrow 0,\ K \leftarrow 0,\ \exists(EM(y,x) \neq 0) \Rightarrow \\ \Rightarrow \exists(EM(y+j,x+i) = 0,\ MT(y+j,x+i) = 0,\ j = \overline{-1,1},\ i = \overline{-1,1}) \Rightarrow \\ \Rightarrow \begin{cases} MT(y+j,x+i) = EM(y,x) \\ K \leftarrow K+1,\ SY(K) \leftarrow y+j,\ SX(K) \leftarrow x+i \end{cases} \end{gathered} \quad (5)$$

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

2.3.2. Осуществляется присоединение полученных соседей к областям экстремумов с помощью выражения

$$\exists(K > 0) \Rightarrow \begin{cases} Flag \leftarrow 1 \\ EM[SY(k),SX(k)] = MT[SY(k),SX(k)] \\ k = \overline{1,K} \end{cases} \quad (6)$$

2.4. Окончание цикла сегментации. Алгоритм выполняется до тех пор, пока не будет найден последний сосед. Если указатель флага $Flag$ после пункта 2.3.2 отличается от нуля, то осуществляется переход в пункт 2.3.1. Иначе, осуществляется выход из алгоритма.



а *б* *в*

Рис. 2. Работа алгоритма сегментации: *а* – экстремумы изображения;
б – добавление пикселей к экстремумам после трех итераций;
в – дерево выращивания 4-связных пикселей *N*-го экстремума

В результате выполнения данного алгоритма формируется матрица сегментации, значение каждого элемента которой указывает на номер сегмента, которому принадлежит пиксель сегментируемого изображения с соответствующими координатами. С каждым циклом размеры сегментов постепенно увеличиваются, в чем проявляется прогрессивный характер сегментации. Число полученных сегментов совпадает с числом локальных экстремумов изображения, что позволяет контролировать это число, а также местоположение и размеры сегментов.

Блок-схема предлагаемого алгоритма представлена на рис. 3.

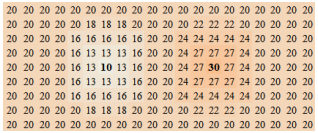
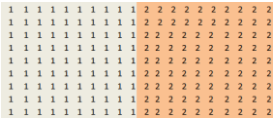
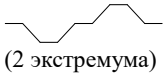
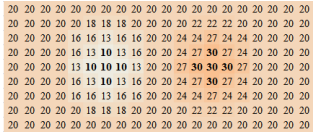
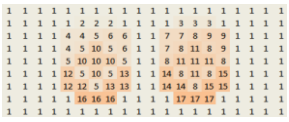
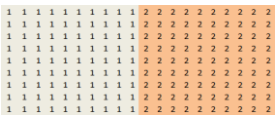
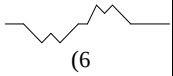
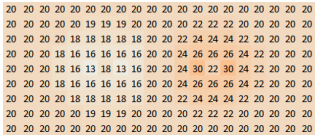
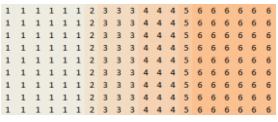
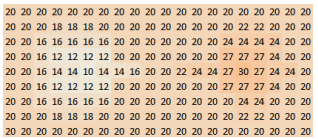
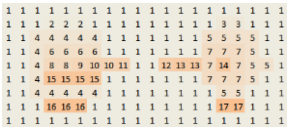
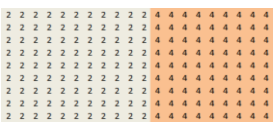
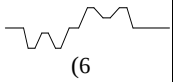
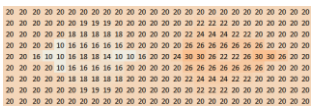
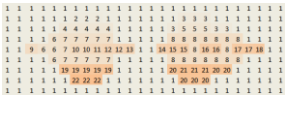
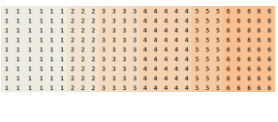
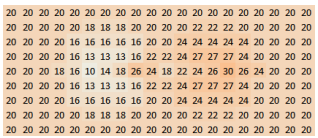
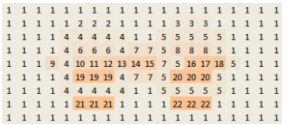
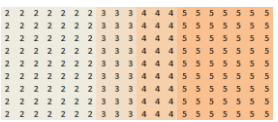
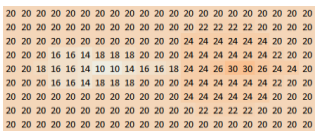
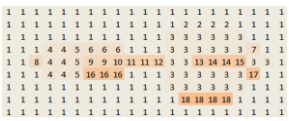
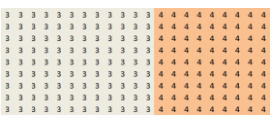
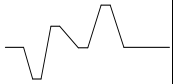
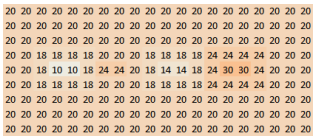
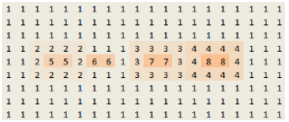
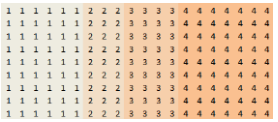
Рис. 3. Блок-схема алгоритма встречного волнового выращивания областей локальных экстремумов

Оценка эффективности алгоритмов сегментации

Выполнена оценка эффективности предложенного алгоритма встречного волнового выращивания областей локальных экстремумов и известного алгоритма выращивания областей. В качестве показателей эффективности алгоритмов использованы время сегментации, возможность выделения объектов, параметризации и компактного описания сегментов.

В табл. 1 приведена визуальная оценка возможности выделения объектов с помощью рассматриваемых алгоритмов сегментации в зависимости от видов экстремумов. Число экстремальных областей экспериментально установлено для тестовых полутоновых изображений, полученных для алгоритма выращивания областей и предложенного алгоритма.

Таблица 1. **Визуальная оценка алгоритмов сегментации в зависимости от различных видов экстремумов**

№	Тестовые изображения		Матрицы сегментации	
	Вид экстремумов	Матрицы изображений	Выращивания областей	Предложенного метода
1	 (2 экстремума)			
2	 (2 экстремума)			
3	 (6 экстремумов)			
4	 (2 экстремума)			
5	 (6 экстремумов)			
6	 (4 экстремума)			
7	 (2 экстремума)			
8	 (4 экстремума)			

Из табл. 1 следует, что предложенный алгоритм выделяет объекты на изображении в зависимости от количества и видов экстремумов. Из матриц сегментации видно, что формы полученных сегментов не являются сложными и зависят от местоположений, форм и расстояний между экстремумами на изображении, что позволяет компактно их описать для последующей обработки. Такое компактное описание сегментов представляет интерес для распознавания и сжатия изображений.

В табл. 2 приведены результаты оценки времени сегментации для рассматриваемых алгоритмов в среде Matlab и Matlab ®Image Processing Toolbox (R2015a), в системе Intel Core i3 3.1 ГГц с 6 ГБ ОЗУ. Для этого эксперимента использованы четыре изображения размерами 128×128, 256×256, 512×512 и 1024×1024 пикселей.

Таблица 2. **Время сегментации тестовых изображений, с**

Алгоритмы сегментации	Тестовые изображения			
	Lena 128×128	Barbara 256×256	France 512×512	Man 1024×1024
Предложенный	0,054706	0,240360	0,657972	4,973419
Выращивания областей	0,137427	0,541161	2,161375	8,739986

Из табл. 2 следует, что предложенный алгоритм выигрывает до 2,5 раз в скорости сегментации по сравнению с алгоиртмом выращивания областей при размере изображения 128×128 пикселей, до 2,3 раз при размере изображений 256×256, до 3,3 раз при размере изображений 512×512 и до 1,8 раз при размере изображений 1024×1024 пикселей.

Заключение

Для сегментации изображений предложен алгоритм на основе встречного волнового выращивания областей локальных экстремумов. Сущность алгоритма заключается во встречном квазипараллельном выращивании областей вокруг локальных минимумов и максимумов, выбранных в качестве начальных точек роста. Это обеспечивает автоматическое разделение областей с плавным перепадом яркости, которые известные алгоритмы сегментируют с ошибками. Показано, что предложенный алгоритм позволяет четко выделять сегменты и контролировать их количество по сравнению с алгоритмом выращивания областей. Установлено, что предложенный алгоритм выигрывает в скорости сегментации до 2,5, 2,3, 3,3 и 1,8 раз по сравнению с алгоритмом выращивания областей при размерах изображений 128×128, 256×256, 512×512 и 1024×1024 пикселей соответственно.

COUNTER-WAVE GROWING OF LOCAL EXTREME REGIONS OF GRAYSCALE IMAGES

NGUYEN ANH TUAN, V.Yu. TSVIATKOU

Abstract. An algorithm for segmentation of halftone images based on counter-wave growth of local extremum regions (minima and maxima) is proposed. In contrast to the known segmentation algorithms, the proposed algorithm allows to separate regions with smooth brightness variations, adapt to a limited segmentation time, and control the number of segments.

Keywords: local extrema, segmentation, wave region growing.

Список литературы

1. Lalitha M., Kiruthiga M., Loganathan C. // J. of Science and Research (IJSR). 2013. Vol. 2. № 2. P. 348–358.
2. Gauch J.M. // IEEE transactions on image processing. 1999. Vol. 8. № 1. P. 69–79.
3. Khiyal M.S.H., Khan A., Bibi A. // Informing Science and Information Technology. 2009. Vol. 6. P. 876–886.
4. Arindrajit S., Arunava D., Prasad S. // International Journal of Computer Science and Information Technologies (IJCSIT). 2015. Vol. 6. № 3. P. 2295–2297.
5. Chang J.H., Fan K.Ch., Chang Y.L. // Image and Vision Computing. 2002. Vol. 20. P. 203–216.
6. Muhsin Z.F., [et. al.] // The Imaging Science Journal. 2014. Vol. 62. № 1. P. 56–62.
7. Xiaolin Wu. // IEEE transactions on pattern analysis and machine intelligence. 1993. Vol. 15. № 8. P. 808–815.
8. Dass R., Priyanka, Devi S. // Int. J. of Electronics & Communication Technology (IJECT). 2012. Vol. 3. Iss. 1. P. 66–70.
9. Singh K.K., Singh A. // Int. J. of Computer Science Issues. 2010. Vol. 7. № 5. P. 414–417.
10. Shih F.Y., Cheng S. // Image and Vision Computing. 2005. № 23. P. 877–886.
11. Sharma R., Sharma R. // Inter. J. of Innovative Research in Computer and Communication Engineering. 2014. Vol. 2. Iss. 9. P. 5686–5692.
12. Mohd Saad N., [et. al.] // Proceedings of the International MultiConference of Engineers and Computer Scientists. Hong Kong, 14–16 March 2012. P. 674–677.
13. Альмияхи О.М., Конопелько В.К. // Матер. междунар. науч.-техн. семинара «Телекоммуникации: сети и технологии, алгебраическое кодирование и безопасность данных». Минск, БГУИР. 2015, С. 61–67.

14. Альмияхи О.М., Цветков В.Ю., Конопелько В.К. // Докл. БГУИР. 2016. № 3 (97). С. 24–30.
15. Рабцевич В.В., Цветков В.Ю., Ловецкий М. Ю. // Сб. матер. междунар. научн.-техн. конф. «Мониторинг техногенных и природных объектов» Минск, БГУИР, 2017. С. 77–84.
16. Рабцевич В.В., Нгуен А.Т., Цветков В.Ю. // Collection of materials of the fourth international scientific and practical conference «BIG DATA Advanced Analytics». Minsk, BSUIR, 2018. P. 373–377.

УДК 528.854

ЛИНЕЙНАЯ НОРМАЛИЗАЦИЯ ЯРКОСТИ АСМ-ИЗОБРАЖЕНИЙ ВЫРАВНИВАНИЕМ ВЫСОТ ОБЪЕКТОВ

В.В. РАБЦЕВИЧ[1], В.Ю. ЦВЕТКОВ[1], А.В. ШАБЛЮК[2]

[1]*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь*
[2]*Институт тепло- и массообмена имени А.В. Лыкова, Республика Беларусь*

Поступила в редакцию 20 марта 2019

Аннотация. Предложен алгоритм линейной нормализации яркости полутоновых изображений атомно-силовой микроскопии выравниванием высот объектов, расположенных на поверхности. Проведено сравнение разработанного алгоритма с алгоритмами вычитания плоскости первого и второго порядков с помощью анализа трехмерных моделей поверхности.

Ключевые слова: нормализация яркости, выравнивание яркости, атомно-силовая микроскопия, вычитание плоскости первого и второго порядков.

Введение

Изображения, получаемые с помощью атомно-силовой микроскопии (АСМ-изображений), часто содержат различные искажения, приводящие к недостоверности получаемых данных о поверхности материалов. Одним из наиболее распространенных видов искажений является наклон поверхности. Он может появиться из-за неточной установки образца, температурного дрейфа или нелинейности перемещений пьезосканера [1].

Целью работы являлась модификация алгоритма нормализации яркостной составляющей изображения, в основе которого лежит нахождение и устранение разницы высот между наиболее схожими областями, и анализ его эффективности. Достоинства алгоритма заключаются в его линейности и учете морфологии изображения. Будет выполнено сравнение с распространенными алгоритмами выравнивания яркости, применяемыми при обработке АСМ-изображений, такими как вычитание плоскости первого и второго порядков [2]. Их реализации представлены в бесплатно распространяемой программе Gwyddion [3].

Алгоритмы выравнивания яркости для АСМ-изображений

Модификация алгоритма добавляет в исходный алгоритм выравнивая яркости АСМ-изображений по высотам их равноразмерных объектов [4] новые шаги для оптимизации получаемого результата. Исходные данные для разработанного авторами алгоритма – полутоновые изображения $F = \left\| f(y,x) \right\|_{(y=\overline{0,Y-1},\, x=\overline{0,X-1})}$, где $f(y,x) \in [0,255]$. Алгоритм состоит из следующих шагов:

Шаг 1. Сегментация исходного изображения F и получение матрицы сегментации S. Следует отметить, что выбор исходного алгоритма сегментации значительно влияет на конечный результат. Например, если ему свойственна недостаточная сегментация или получаемые области не коррелируют со значениями яркости на исходном изображении, необходимые данные для дальнейшей нормализации не будут получены. В данной реализации используется алгоритм регрессивного волнового выращивания областей АСМ-изображений с автоматическим обнаружением границ [5].

Шаг 2. Деление матрицы S на четыре равные части C_1, C_2, C_3, C_4.

Шаг 3. Нахождение в каждой области C_1, C_2, C_3, C_4 наиболее равноразмерных сегментов g_1, g_2, g_3, g_4, каждый из которых полностью принадлежит одной из областей.

Шаг 4. Нахождение самого яркого P_{max} и самого тусклого P_{min} пикселей, которые будут принадлежать сегментам g_1, g_2, g_3, g_4.

Шаг 5. Нахождение приращения Δ на каждом шаге $\Delta = \frac{P_{max} - P_{min}}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$, где x_1, x_2 и y_1, y_2 – координаты точек P_{max} и P_{min}.

Шаг 7. Определение ориентации в пространстве друг относительно друга двух областей с максимальной (P_{max}) и минимальной (P_{min}) яркостью. В алгоритме возможны три варианта ориентации: вертикальная, горизонтальная и диагональная. В изображениях со сложной морфологией возможны как автоматическое определение расположение областей, так и статическая установка необходимой ориентации.

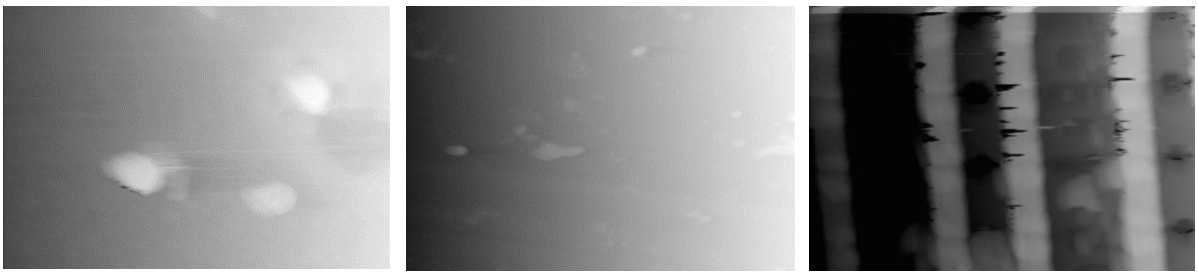
Шаг 8. Определение точек, принадлежащих линии раздела, как средних между координатами центров двух сегментов. Расположение линии раздела на плоскости будет зависеть от ориентации, найденной ранее.

Шаг 9. Добавление или исключение на исходном изображении (начиная от линии раздела) $\Delta \times i$ при приближении к P_{max} или P_{min} соответственно. В результате работы алгоритма формируется новое изображение с выровненной яркостью. Алгоритм позволяет внести изменения в изображение, не нарушая его структуры [4].

Оценка эффективности алгоритмов выравнивания яркости

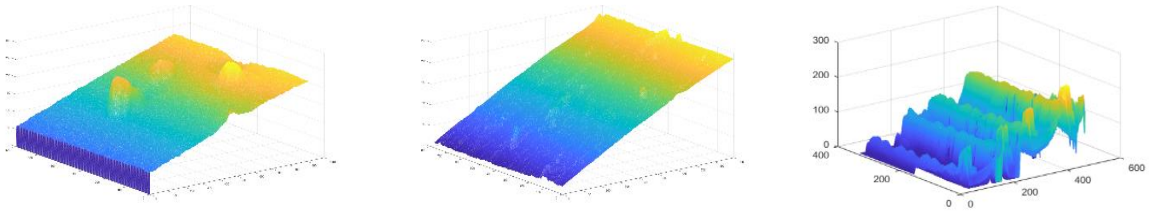
Для анализа использованы следующие алгоритмы, часто применяемые для устранения наклона образца: вычитание плоскости первого и второго порядков (значения пикселей изображения аппроксимируются полиномом заданного порядка, после этого выделяется и удаляется составляющая фона) [1].

Для оценки эффективности работы алгоритмов нормализации яркости были подобраны АСМ-изображения с заметным наклоном образца (рис. 1) и построены трехмерные модели их поверхностей (рис. 2). На рис. 3–5 представлены результаты нормализации яркости разработанным алгоритмом и при вычитании плоскости первого и второго порядков.



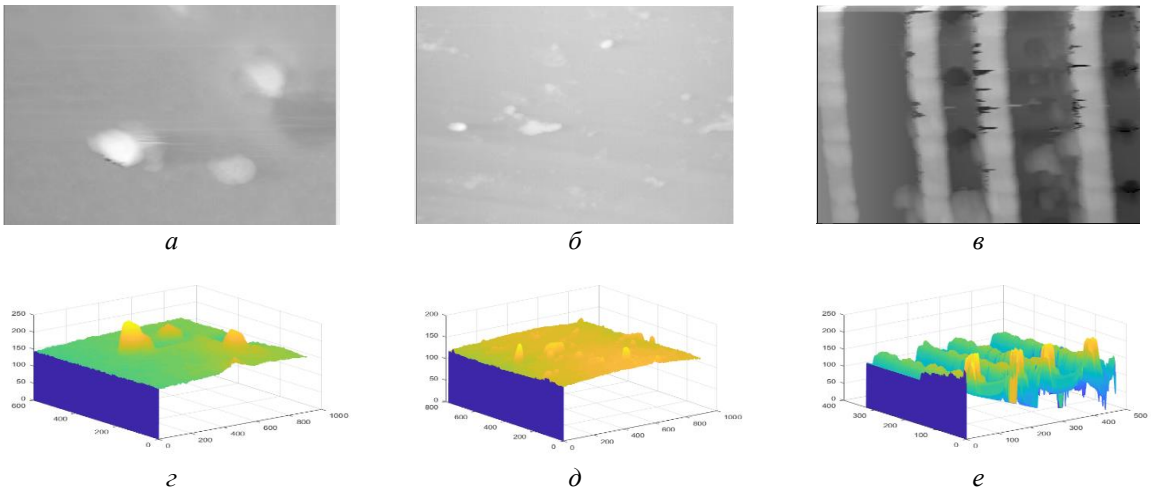
а *б* *в*

Рис. 1. Исходные АСМ-изображения: *а* – АСМ 1; *б* – АСМ 2; *в* – АСМ 3



а *б* *в*

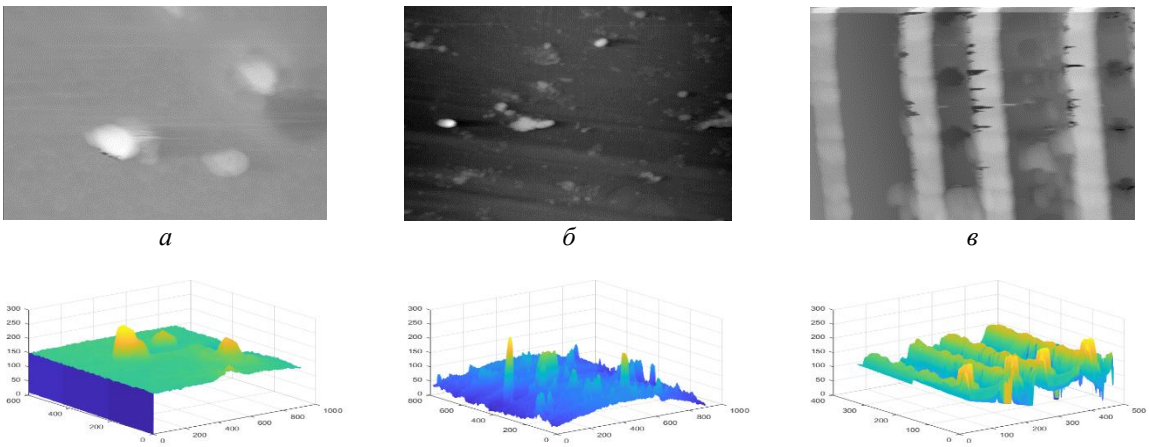
Рис. 2. Трехмерные модели поверхностей изображений 1 (*а*), 2 (*б*), 3 (*в*)



а *б* *в*

г *д* *е*

Рис. 3. Результаты нормализации яркости с помощью разработанных алгоритмов (*а, б, в*); и трехмерные модели поверхностей образца (*г, д, е*)

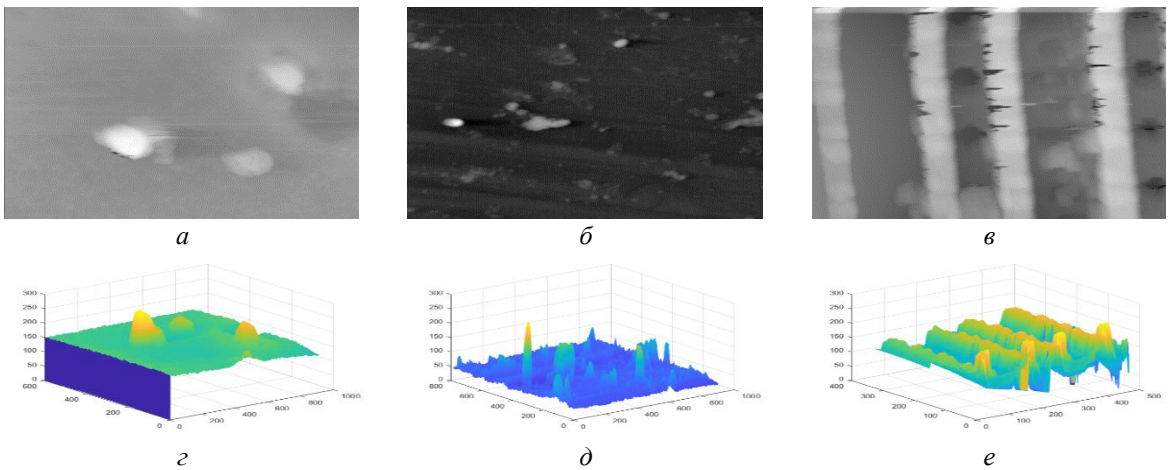


а *б* *в*

г *д* *е*

Рис. 4. Результат нормализации яркости при вычитании плоскости первого порядка (*а, б, в*); и трехмерные модели поверхностей образца (*г, д, е*)

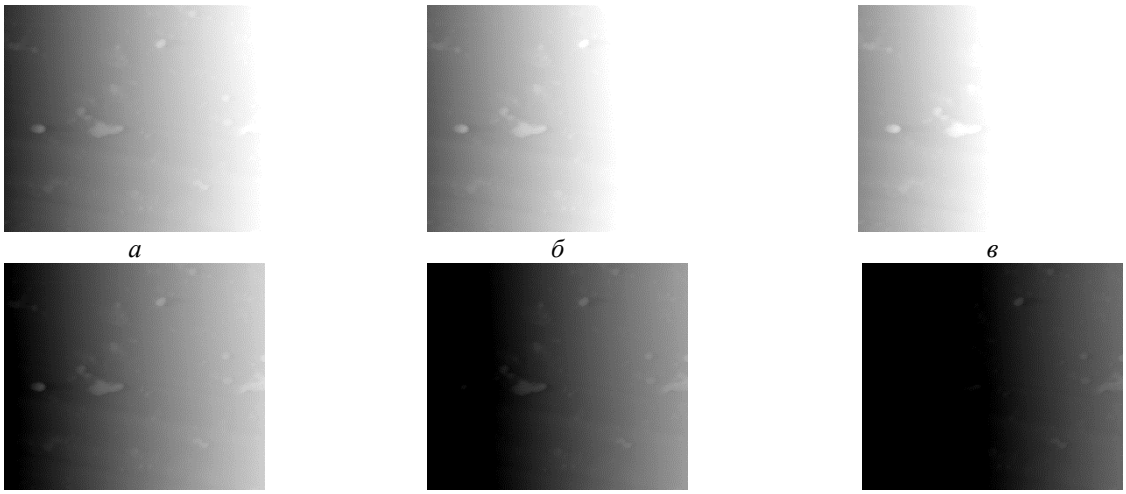
Для оценки результатов нормализации яркости было выбрано контрольное тестовое изображение и выполенено увеличение и уменьшение его яркости на 10, 30 и 50 процентов (рис. 6), построены трехмерные модели поверхностей (рис. 7). На рис. 8–10 представлены результаты нормализации яркости контрольного изображения разработанным алгоритмом и при вычитании плоскости первого и второго порядка, соответственно.



а *б* *в*

г *д* *е*

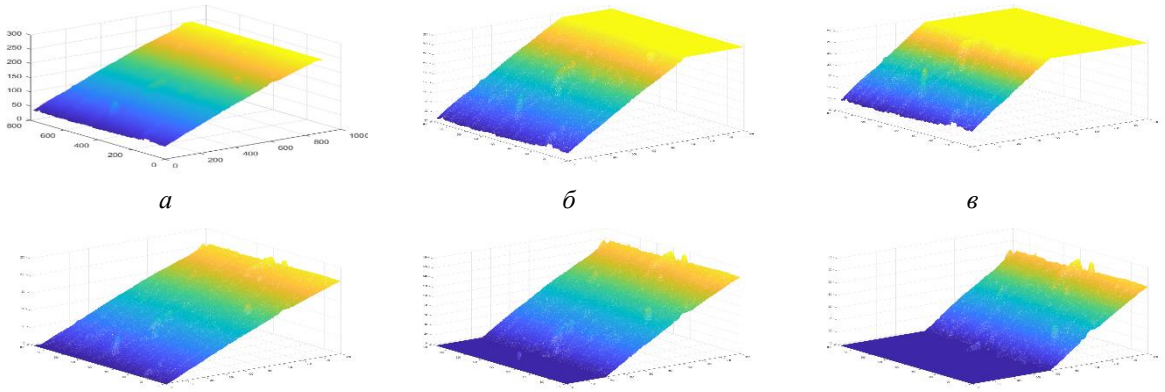
Рис. 5. Результат нормализации яркости при вычитании плоскости второго порядка (*а, б, в*); и трехмерные модели поверхностей образца (*г, д, е*)



а *б* *в*

г *д* *е*

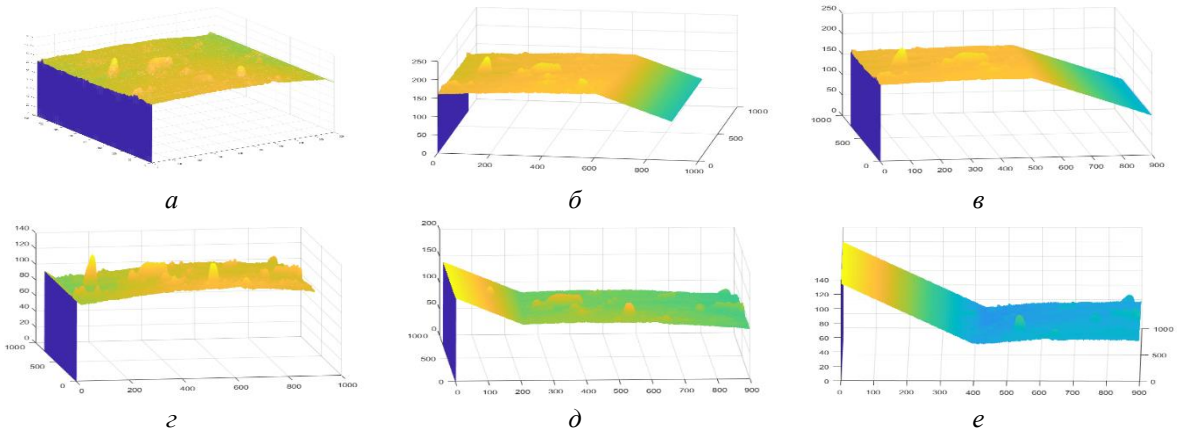
Рис. 6. Контрольное изображение с увеличенной на 10 % (*а*), 30 % (*б*), 50 % (*в*), и уменьшенной на 10 % (*г*), 30 % (*д*), 50 % (*е*) яркостью



а *б* *в*

г *д* *е*

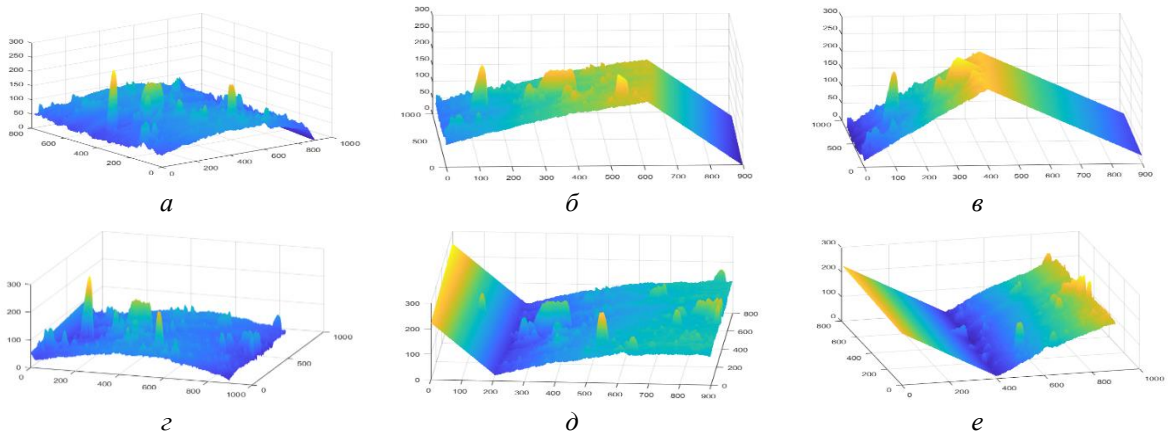
Рис. 7. Трехмерная поверхность контрольного изображения с увеличенной на 10 % (*а*), 30 % (*б*), 50 % (*в*), и уменьшенной на 10 % (*г*), 30 % (*д*), 50 % (*е*) яркостью



а *б* *в*

г *д* *е*

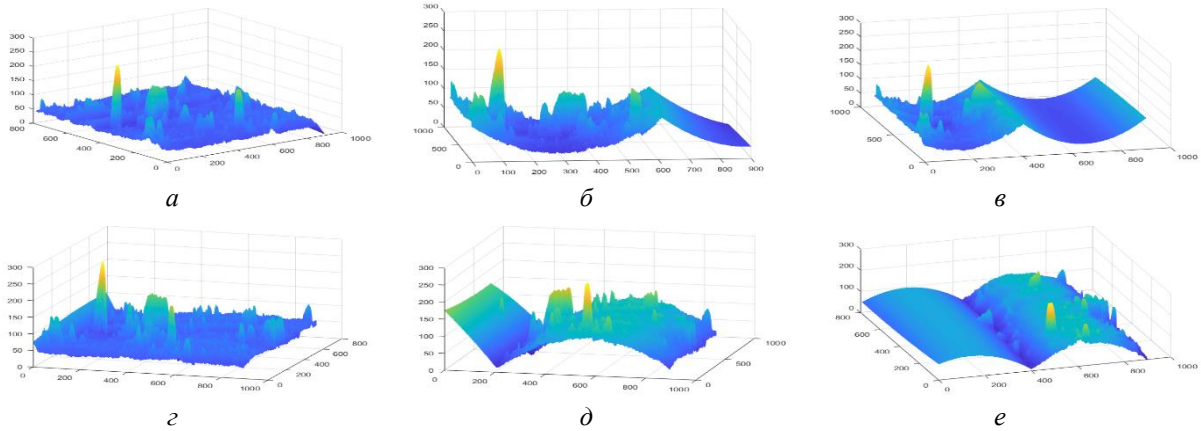
Рис. 8. Трехмерные поверхности контрольного изображения с увеличенной на 10 % (*а*), 30 % (*б*), 50 % (*в*), и уменьшенной на 10 % (*г*), 30 % (*д*), 50 % (*е*) яркостью после нормализации яркости разработанным алгоритмом



а *б* *в*

г *д* *е*

Рис. 9. Трехмерные поверхности контрольного изображения с увеличенной на 10 % (*а*), 30 % (*б*), 50 % (*в*), и уменьшенной на 10 % (*г*), 30 % (*д*), 50 % (*е*) яркостью после вычитания плоскости первого порядка



а *б* *в*

г *д* *е*

Рис. 10. Трехмерные поверхности контрольного изображения с увеличенной на 10 % (*а*), 30 % (*б*), 50 % (*в*), и уменьшенной на 10 % (*г*), 30 % (*д*), 50 % (*е*) яркостью после вычитания плоскости второго порядка

На основе анализа представленных трехмерных поверхностей можно сделать вывод о том, что все алгоритмы способны убрать наклон и нормализировать поверхность, однако при внесении дополнительных яркостных искажений только разработанный алгоритм сохраняет линейность формы трехмерной поверхности изображения. Остальные методы способны внести дополнительные нелинейные искажения, которые, в свою очередь, могут привести к искривлению границ частиц, расположенных на поверхности.

Заключение

Результат работы созданного алгоритма сильно зависит от метода сегментации, выбранного на начальном этапе. При недостаточной сегментации исходного изображения будет невозможно выполнить все последующие шаги нормализации

На основе анализа представленных трехмерных поверхностей можно сделать вывод о том, что использование алгоритмов вычитания плоскости первого и второго порядков может оставить искривление поверхности и исказить границы исходных объектов. Применение разработанного алгоритма линейной нормализации не меняет форму поверхности при внесении дополнительных искажений.

LINEAR BRIGHTNESS NORMALIZATION OF AFM-IMAGES BY HEIGHT LEVELLING OF OBJECTS

V.V. RABTSEVICH, V.Yu. TSVIATKOU, A.V. SHABLIUK

Abstract. The algorithm for half-tone images of atomic force microscopy linear brightness normalization, by leveling the heights of objects located on the surface, is proposed. The developed algorithm is compared with the first and second order levelling algorithms using the analysis of three-dimensional models of the surface.

Keywords: brightness normalization, brightness levelling, atomic force microscopy, subtraction of a plane, the first and second order, levelling.

Список литературы

1. Миронов В.Л. Основы сканирующей зондовой микроскопии Нижний Новгород: РАН Институт физики микроструктур, 2004.
2. Eaton P., West P. Atomic Force Microscopy London: Oxford University Press, 2011. P. 105–108.
3. Gwiddion [Электронный ресурс]. URL: http://www.gwiddion.net (дата обращения: 02.03.2019).
4. Рабцевич В.В., Цветков В.Ю., Ловецкий М.Ю. // Матер. XVII Междунар. конф. «Развитие информатизации и государственной системы научно-технической информации» Минск, 20 сентября 2018 г. С. 190–195.
5. Рабцевич В.В., Цветков В.Ю. // Сб. материалов V Междунар. науч.-практ. конф. «BIG DATA and Advanced Analytics» Минск, 13–14 марта 2019 года В 2 ч. Ч.2. Минск, БГУИР, 2019. С. 203–209

UDC 004.932.72'1;004.93'14

A NEW IMPROVED FAST PARALLEL SKELETONIZE ALGORITHM

Ma JUN, V.Yu. TSVIATKOU, V.K. KONOPELKO

Belarusian state university of informatics and radioelectronics, Republic of Belarus

Submitted 8 March 2019

Abstract. The skeletonize of the binary image has crucial application in the field of target recognition. The thinning result of the Zhang's fast parallel thinning algorithm maintains the connection of the original image, has a good structural form and has no burs. However, Zhang's fast parallel thinning algorithm consists two subiterations in where the most of operations are similar, which cause the waste of the resources of the calculate. Besides, it can't ensure the single pixel width, which brings difficulties for post-processing. In this paper, authors make some improvement on the Zhang's fast thinning algorithm. Experimental results show that the use of improved algorithm can not only reduce the total number of the iteration but also ensure a single pixel width.

Keywords: Zhang's fast parallel thinning algorithm, number of iterations, single pixel

Introduction

Skeletonization, also known as the image thinning, is a pre-processing which is widely applied in field of the image processing and pattern recognition. Image thinning refers to finding the skeleton or centerline of the original image as quickly as possible while maintaining the completeness of the topology of the original image, and then replacing the original image with a single-pixel skeleton. The processing of the image thinning can dramatically reduce the superfluous information from the image, which relieve the computation burden of the computer and shorten the time which spend on the process of the recognition [1, 2].

Skeletonization is divided into two main approaches: iterative and non-iterative. In non-iterative approach, the skeleton is extracted directly without examine each pixel individually, however these methods are difficult to implement and slow as well. As for iterative techniques, which can further divided into two methods: the method of parallel processing and the method of processing sequentially. In the parallel way the whole unwanted pixels are erased after identify the whole wanted pixels, whereas in sequential way; the unwanted pixels are removed in the identifying the desired pixels in each iterative [3].

Among the parallel thinning algorithms, Zhang's algorithm [1] has the good performance in continuity and it can relatively precise describe the straight line, inflection point and cross point. However, at the same time, Zhang's algorithm also has some aspects can improved, which will have better performance, this paper will base on Zhang's algorithm to propose a improved ones.

Zhang's Algorithm

Zhang's algorithm consists of two sub-iterations. Iterative transformations are applied to original binary matrix point by point according to the values of a small set of neighboring points. It is assumed that a 3×3 window is used, and that each element connected with its eight neighboring elements which is shown in Fig. 1 [1, 4].

P_9 $(i-1, j-1)$	P_2 $(i-1, j)$	P_3 $(i-1, j+1)$
P_8 $(i, j-1)$	P_1 (i, j)	P_4 $(i, j+1)$
P_7 $(i+1, j-1)$	P_6 $(i+1, j)$	P_5 $(i+1, j+1)$

Fig. 1. Designations of the nine pixels in a 3×3 window

In the first sub-iteration, the contour point P_1 is deleted from the digital pattern if it satisfies the following conditions:

$(a)\ \ 2 \le B(P_1) \le 6,$

$(b)\ \ \mathrm{A}(\mathrm{P}_1) = 1,$

$(c)\ \ P_2 \cdot P_4 \cdot P_6 = 0,$

$(d)\ P_4 \cdot P_6 \cdot P_8 = 0,$

where $A(P_1)$ is the number of 01 patterns in the ordered set $P_2, P_3, P_4, ...P_8, P_9$ which are the eight neighbors of P_1.

$B(P_1)$ is the number of nonzero neighbors of P_1,which means that

$$B(P_1) = P_2 + P_3 + P_4 + \cdots + P_8 + P_9.$$

In the second sub-iteration, only conditions (c) and (d) are changed as follows:

$(c')\ \ P_2 \cdot P_4 \cdot P_8 = 0,$

$(d')\ P_2 \cdot P_6 \cdot P_8 = 0.$

Zhang's algorithm works by following way. First, finding points which are satisfied the conditions of the first sub-iteration and delete them, and then location these points which are satisfied the conditions of the second sub-iteration and deleting them. The algorithm will not stop and continued to repeat this order forever until there is not any point which can be deleted.

The above algorithm has very good results with respect to both connectivity and contour noise immunity. Furthermore, the conditions for searching those points that should be deleted from the pattern are very simple.

However, the main iteration of the Zhang's algorithm consists of two sub-iteration which means it needs two full scans of the image to eliminate the contour points, which cause the costs of the time is relatively high.

Moreover, Zhang's algorithm has not presented an ideal performance in single pixel, which also seen as the main factor to evaluate a thinning algorithm.

So, the proposed method is aimed to tackle the problem which mentioned above. This method will reduce the number of the total iterations and, at the same time, realize the single pixel as far as possible.

Improvement Algorithm

Improvement algorithm include 4 main parts: search module, connectivity check module, single pixel correction module and contour point delete module. The structure of these stages addressed in Flowchart as shown in Fig. 2.

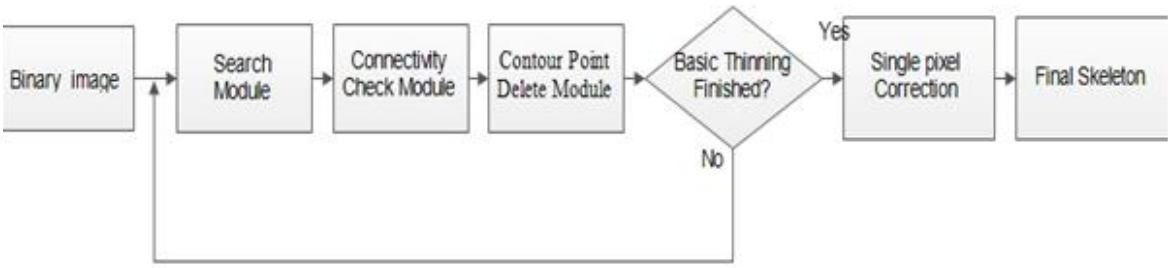


Fig. 2. Flowchart of the purposed method

The binary image is acquisition into the proposed method as black pixels which considered as foreground as well as consider an object pixel for deletion. The pixels having value 0 are considered as background pixels.

Before to introduce the search module, the 3×3 window which we used in the search module need to present, which is a little different from Zhang's algorithm as shown in Fig. 3.

P_5	P_4	P_3
P_6	P_1	P_2
P_7	P_8	P_9

Fig. 3. Designation of the 3×3 window

In the search module, 3×3 window is used to scan the point by point of the original binary image. The center of the window aims to the current pixel. If the current pixel is a foreground point, it's further observed its neighbor is match the corresponding conditions or not. And then if this point satisfied the conditions, we will consider it as a potential contour point and record it as «1» in the matrix *S*. But in order to give the condition, we divided all foreground points into two groups. One is named as general points set, another is named as special points set. The conditions of the general points set are completely the same as the first two condition (condition (*a*) and (*b*)) from Zhang's algorithm. The special points are 4 points which have shown below, they are «0» appear in the upper left corner (Fig. 4, *a*), in the upper right corner (Fig. 4, *b*), in the lower left corner (Fig. 4, *c*), in the lower right corner (Fig. 4, *d*) respectively.

0	1	1		1	1	0		1	1	1		1	1	1	
1	1	1		1	1	1		1	1	1		1	1	1	
1	1	1		1	1	1		0	1	1		1	1	0	
a				*b*				*c*				*d*			

Fig. 4. All Special Points

When faced with a special point which is looked like the one of Fig. 4, it's necessary to know these points which are sat around the 0 in this window are potential contour points or not. If they are all potential contour points, the current pixel is also a potential contour point. Otherwise, the current pixel cannot consider as a potential contour point. For example, if we meet a point which is looked like as (*b*) (Fig. 4). We need to confirm the P_4 and P_2 (Fig. 3). We know that P_4 has been checked because the location of P_4 is prior to the current pixel, as a result, we can straight take out the value of $S(x4, y4)$, where $x4$, $y4$ is the coordinate point of the P_4 to see P_4 is a potential contour point or not.

And then we need to check P_2, we didn't know the value of the $S(P_2)$, because the location of P_2 is behind current pixel, so we have to turn our 3×3 window to this pixel and test its neighbor to see it is a potential contour point or not. The judgment of the P_2 is as the same as the general points. When P_2 and P_4 are potential contour point, the original pixel will be returned. It will be considered as a potential contour point.

The function of the connectivity check module is to verify the connectivity under the hypothesis that we have deleted the potential contour points which stored in the matrix *S*. If the deletion of some points will break the connectivity of the image, this potential contour points will be excluded and it will be written as «0» in the matrix *S*. For different type of point set, deploying the different conditions of judgment is reasonable. If the 8-neighbor of a general point appeared as the same as the retention template which has shown in Fig. 5, it's necessary to retain this point.

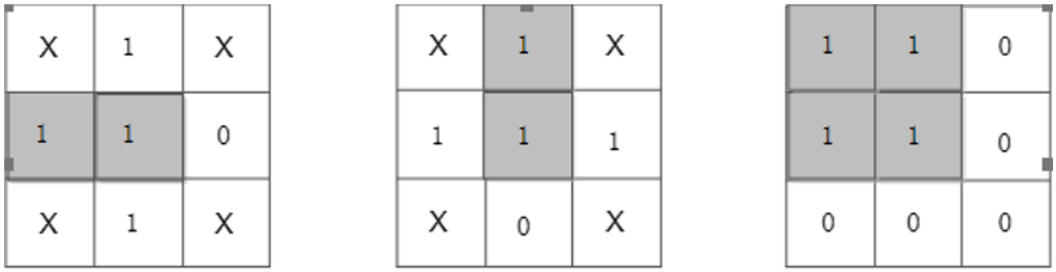


Fig. 5. Retention template for general point

In retention template, the pixel with deep color means that this pixel has been viewed as a potential contour point in the search module.

As for the special point, 4-neighbor points will be observed, if there are more than 3 points have been regarded as potential contour points, the current pixel deserved to keep.

After the connectivity check module, some potential contour points which may break the connectivity of the binary picture are eliminated. And then it's possible to remove contour points according to the matrix *S* from the original binary image with the confidence.

The last step is to conduct a single pixel correction, which is based on the result of the previous image thinning procession. After referred to the [4], a single pixel result is got if these points according to the correction template are deleted (Fig. 5).

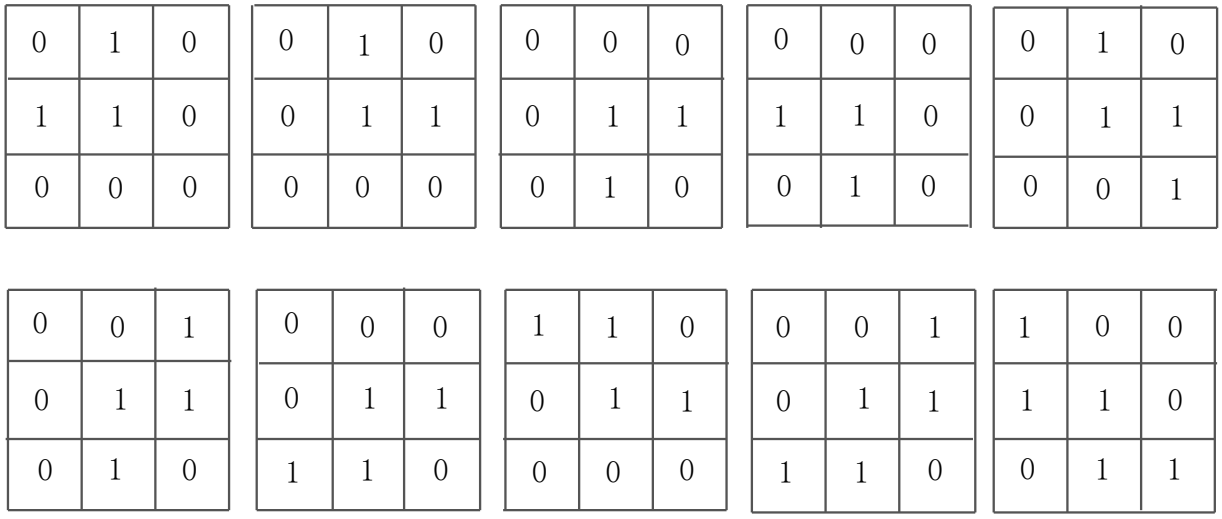


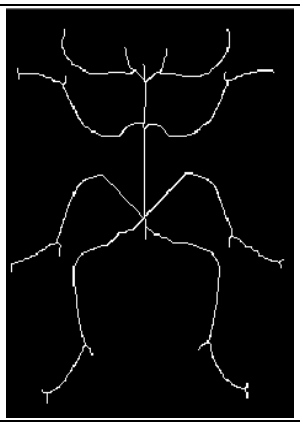
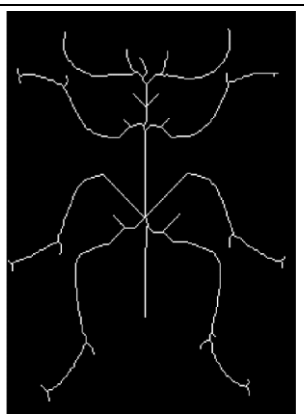
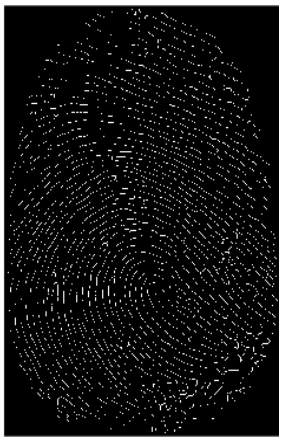
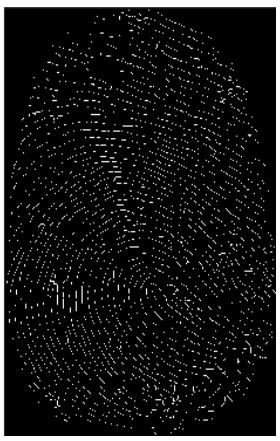
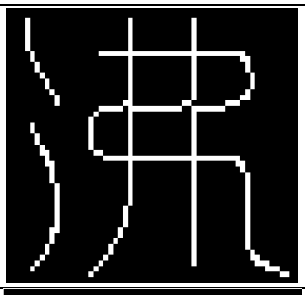
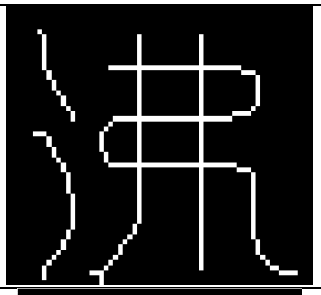
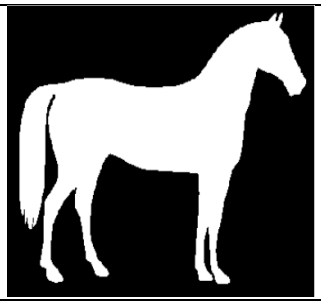
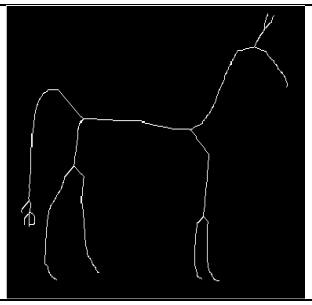
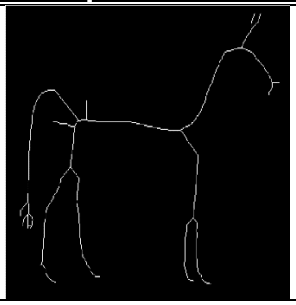
Fig. 6. Single pixel correction template

In order to easy to identify these points, it's possible to translate it into the binary code. The sequence is ordered according the 3×3 window which has shown in Fig. 1. The most significant bit is P_9 and least signification bit is P_2.

Experiments and results

To assess the performance, the improvement algorithm and the Zhang's algorithm were written in Matlab R2018b.This data set has many class shapes. The achieved results are presented in Tables 1 and 2.

Table 1**. The original binary images and the thinning results by different algorithm**

	Original binary image	Zhang's algorithm	Improved algorithm
a			
b			
c			
d			
e			

According to the Table 1, new method has a better performance in the single pixel.

Table 2. **Comparison of the compute speed between algorithms**

Original binary image	Number of iterations		CPU time consumed(s)	
	Zhang's algorithm	Improved algorithm	Zhang's algorithm	Improved algorithm
a	14	8	0,3074	0,1846
b	82	42	0,6683	0,4427
c	63	32	9,8992	6,0417
d	18	9	0,1137	0,0757
e	89	45	1,9470	1,0489

Thus, improved algorithm has better performance in the single pixel and in the speed, which thanks to the reduction of the number of the iteration.

Theoretically, because the number of the iteration of the improved algorithm is half of the Zhang's algorithm, The CPU time consumed of our algorithm should also be a half when compare with the Zhang's algorithm. However, the result of the experiments reflect that proposed algorithm only cut down about 35 % time which as shown in the Table 2, because the single pixel is added. Correction parts to the algorithm increases the complex of the compute and prolong the execution time.

However, proposed algorithm has some weak point. With the increasement of speed, some burrs appear in the algorithm results. This problem can be solved by refer to the [5].

Conclusion

In this paper, we presented an improved algorithm based on the zhang's algorithm, which has better performance in speed and single pixel. The experiments have proved the effectivity of the new algorithm.

References

1. Zhang T.Y., Suen C.Y. // Communications of ACM. 1984. Vol. 27, No 3. P. 236–239.
2. Бушенко Д.А., Садыхов Р.Х. // Технология построения алгоритмов утоньшения для скелетизации протяженных объектов. 2009. No 7 (45). P. 81–86.
3. Waleed A.A. [et. al.] // Skeletonization Algorithm for Binary Images. Procedia Technology. 2013. P. 704–709.
4. Mu S.M., Du H.Y., Su P., Zhu X.H., Chen G.Y. // Microelectronics & Computer.2013. Vol. 30, No. 1. P. 53–55.
5. Ye F.L. // J. of Xichuang University. .2018. Vol. 32 (3), P. 91–93

УДК 621.391

ПАРАМЕТРИЗАЦИЯ АСМ-ИЗОБРАЖЕНИЙ НА ОСНОВЕ ТРИАНГУЛЯЦИИ

В.А. КОВШИК, Н.С. ДАВЫДОВА, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Разработан алгоритм разбиения топографического пространства на симплексы на основе триангуляции для параметризации изображений атомно-силовой микроскопии (АСМ-изображений). Алгоритм обеспечивает формирование геометрического описания областей АСМ-изображений, соответствующих элементам поверхностей неорганических материалов. Исследована устойчивость алгоритма к изменению яркости и контраста АСМ-изображения.

Ключевые слова: АСМ-изображения, триангуляция, параметризация.

Введение

Определение параметров структурных областей на изображении является одной из ключевых задач при выявлении зависимостей «структура-свойства». АСМ часто является наиболее информативным методом исследования поверхности и результатов ее модификации в субмикро- и наноразмерном диапазоне.

Идентификационные параметры, которые можно выделить для распознавания и описания областей на изображении, делятся на три группы: собственные, окрестные и относительные [1].

К свойствам значений относят: минимум, (минимальное значение высоты, найденное внутри сегмента), максимальное значение высоты, найденное внутри сегмента, среднее значение всех принадлежащих сегменту пикселей.

К свойствам площадей можно отнести площадь проекции найденного сегмента на плоскость, сторону эквивалентного квадрата (т. е. сторону квадрата с той же площадью проекции, что и у найденного сегмента, радиус эквивалентного диска (радиус диска с той же площадью проекции, что и у найденного сегмента), площадь поверхности (площадь поверхности найденного сегмента).

К свойствам границ относятся длина границы проекции (длина проекции границы сегмента на горизонтальную плоскость, а не на реальную трехмерную поверхность), минимальный ограничивающий размер (минимальный размер сегмента в горизонтальной плоскости), максимальный ограничивающий размер (максимальный размер зерна в горизонтальной плоскости). Максимальные и минимальные ограничивающие размеры зерна показаны на рис. 1 [2].

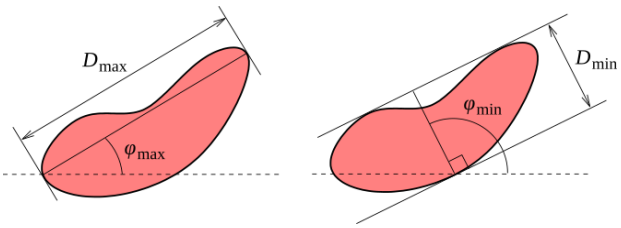


Рис. 1. Максимальные и минимальные ограничивающие размеры зерна

К свойствам объемов можно отнести объем между нулевой плоскостью и поверхностью (объем между поверхностью зерна и плоскостью $Z = 0$), объем между поверхностью и плоскостью, проходящей через минимум (объем между поверхностью зерна и плоскостью

$Z = Z\min$, где $Z\min$ минимальная высота, найденная внутри сегмента).

К свойствам расположения можно отнести положение центра по оси абсцисс (горизонтальная координата центра сегмента), положение центра по оси ординат (вертикальная координата центра сегмента).

К свойствам наклонов относятся наклон θ (отклонение нормали к средней плоскости от оси Z), наклон φ (азимут наклона).

К свойствам кривизны можно отнести центр кривизны по оси абсцисс (горизонтальное положение центра квадратичной поверхности, которой аппроксимируется поверхность зерна), центр кривизны по оси ординат (вертикальное положение центра квадратичной поверхности, которой аппроксимируется поверхность зерна).

К относительным параметрам можно отнести свойства, описывающие место области относительно других выделенных областей на изображении. Одним из методов отображения таких параметров является построение сеток. Например, триангуляция – определение взаимного расположения точек на поверхности при помощи построения сети треугольников [2].

Алгоритм параметризации АСМ-изображений на основе триангуляции

Для параметризации АСМ-изображений предлагается алгоритм разбиения топографического пространства на симплексы на основе триангуляции. Сущность алгоритма состоит в нахождении локальных максимумов в каждой из сегментированных областей, а затем их соединении с двумя ближайшими максимумами. В результате выполнения алгоритма формируется множество треугольников, вершины которых соединяют все сегменты исходного изображения, а также матрицы углов и расстояний, описывающих эти треугольники. Алгоритм состоит из следующих шагов.

1. Загрузка матрицы зондирования, в которой значение каждого элемента $m_Z(y,x) \in \left[0, 2^B - 1\right]$ определяется высотой (трением или вязкостью) соответствующей точки поверхности, где Y, X – размеры матрицы зондирования по вертикали и горизонтали.

2. Формирование матрицы $M_S = \left\| m_S(y,x) \right\|_{\left(y=\overline{0,Y-1},\, x=\overline{0,X-1}\right)}$ сегментации по алгоритму [3], значимые области которой нумеруются с помощью счетчика C_S сегментов.

3. Формирование матрицы $M_Q(C_S) = \left\| m_S(C_S,y,x) \right\|_{\left(y=\overline{0,Y-1},\, x=\overline{0,X-1}\right)}$ значений локальных максимумов для каждого значения C_S, элементы которой определяются с помощью выражения

$$\begin{cases} \left(m_Z(y,x) = \max C_S\right) \Rightarrow \left(m_Q(C_S,y,x) \leftarrow 1\right), \\ \left(m_Z(y,x) \neq \max C_S\right) \Rightarrow \left(m_Q(C_S,y,x) \leftarrow 0\right), \end{cases}$$

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

4. Формирование стека S со значениями координат экстремумов матрицы M_Q.

5. Формирование стеков S_1 и S_2 со значениями двух ближайших к экстремуму рассматриваемого сегмента.

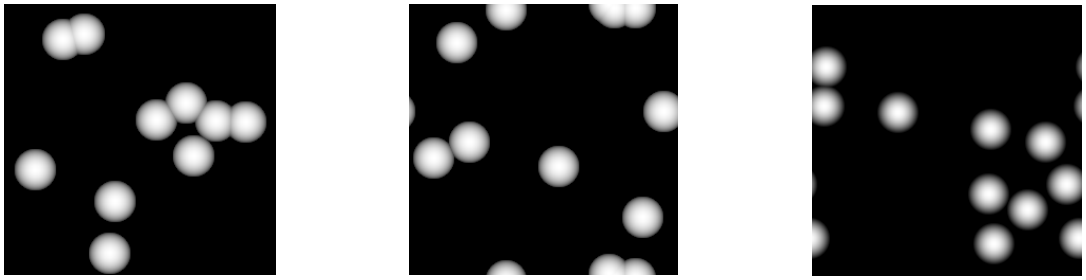
6. Формирование стеков R_1 и R_2 со значениями расстояний к двум ближайшим экстремумам от экстремума рассматриваемого сегмента.

7. Построение матрицы $M_c = \left\| m_c(y,x) \right\|_{\left(y=\overline{0,Y-1},\, x=\overline{0,X-1}\right)}$ с разбиением топографического пространства на симплексы.

Алгоритм параметризации АСМ-изображений на основе триангуляции реализован в среде Matlab.

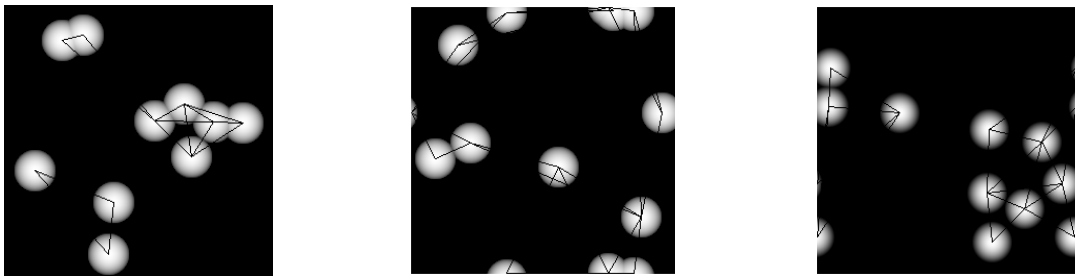
Оценка работы алгоритма параметризации АСМ-изображений на основе триангуляции

На рис. 2 приведены тестовые изображения, а на рис. 3 и 4 – полученные на их основе с помощью разработанного алгоритма триангуляционные сетки.



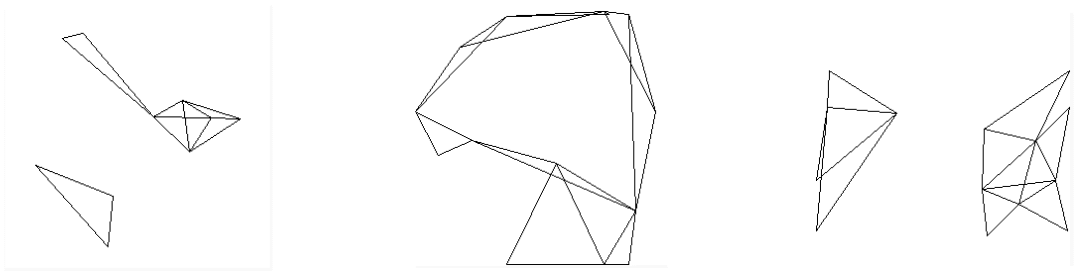
а *б* *в*

Рис. 2. Тестовые АСМ-изображения: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3



а *б* *в*

Рис. 3. Результаты триангуляции тестовых АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3



а *б* *в*

Рис. 4. Полученные сетки тестовых АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3

В таблице для тестового изображения ТР-1 приведены значения углов и расстояний между локальными максимумами областей, полученные с помощью предложенного алгоритма параметризации АСМ-изображений на основе триангуляции.

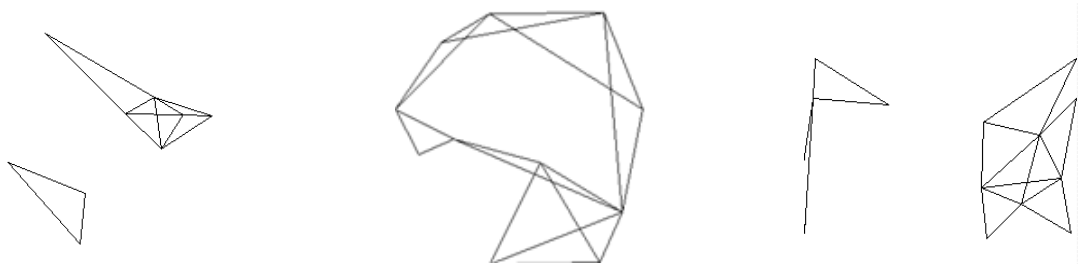
Значения углов и расстояний для найденных центров масс

Углы, град	Расстояние, пиксел	Расстояние, пиксел	Расстояние, пиксел
26,655	80,777	105,551	49,254
54,851	20,616	116,276	105,759
115,977	20,616	105,759	116,276
47,370	49,254	105,551	80,777
105,975	49,254	80,777	105,551
73,914	32,249	48,796	50,488
118,991	32,249	32,757	56,009
78,306	39,115	48,795	56,009
150,782	28,018	32,757	58,822
35,192	28,0179	58,523	39,115

Из таблицы следует, что с помощью предложенного алгоритма можно найти экстремумы всех значимых областей и формирует простые симплексы на основе анализа соседних центров.

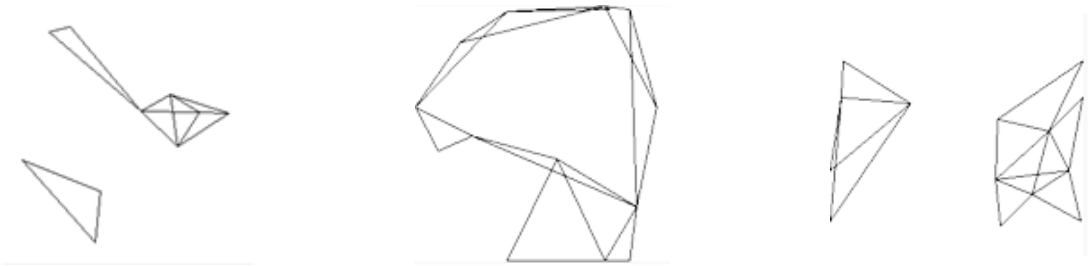
Оценка устойчивости алгоритма параметризации АСМ-изображений на основе триангуляции

Проверка устойчивости алгоритма разбиения топографического пространства на симплексы на основе триангуляции проводилась при изменении параметров яркости и контрастности. Полученные результаты приведены на рис. 5–8.



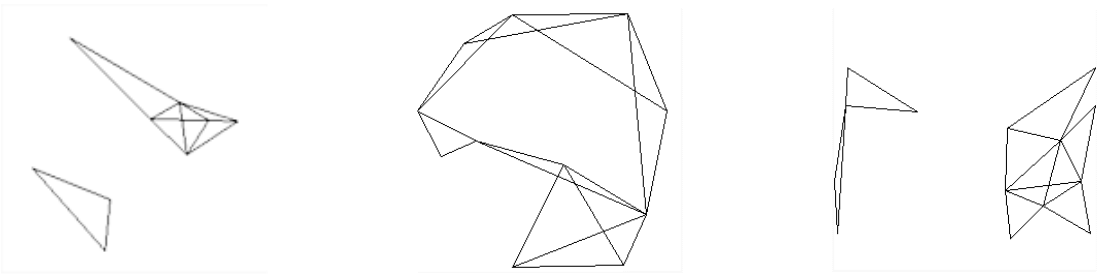
а *б* *в*

Рис. 5. Результат триангуляции при увеличении яркости АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3



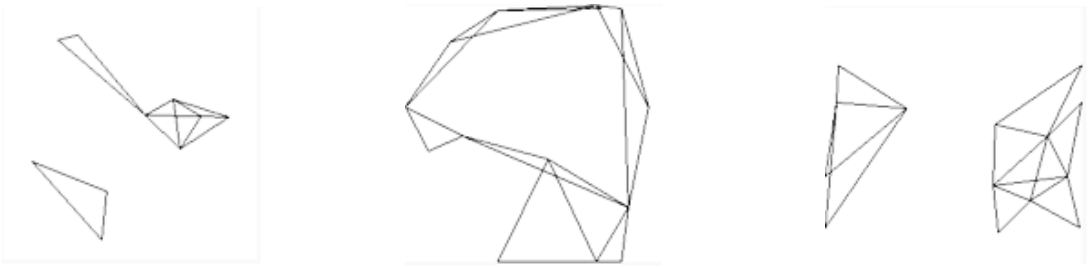
а *б* *в*

Рис. 6. Результат триангуляции при уменьшении яркости АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3



а *б* *в*

Рис. 7. Результат триангуляции при увеличении контрастности АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3



а *б* *в*

Рис. 8. Результат триангуляции при уменьшении контрастности АСМ-изображений: *а* – ТР-1; *б* – ТР-2; *в* – ТР-3

Из рис. 5–8 видно, что уменьшение яркости и контраста АСМ-изображений не влияет на результаты триангуляции, а при увеличении яркости и контраста появляются изменения

из-за засвета границы близко расположенных зерен. Эту проблему можно решить предварительной обработкой АСМ-изображений.

Заключение

Разработан алгоритм разбиения топографического пространства на симплексы на основе триангуляции для параметризации АСМ-изображений. Алгоритм формирует геометрическое описание областей, соответствующих элементам поверхностей неорганических материалов, в виде сетки треугольников. Приведены основные характеристики, которые можно использовать в качестве собственных и относительных идентификационных параметров для распознавания и описания областей на АСМ-изображении. Показано, что разработанный алгоритм устойчив к уменьшению яркости и контраста АСМ-изображений при их предварительной нормализации.

PARAMETRIZATION OF AFM-IMAGES BASED ON TRIANGULATION

V.A. KOVSHIK, N.S. DAVYDOVA, V.Yu. TSVIATKOU

Abstract. An algorithm has been developed for splitting topographic space into simplexes based on triangulation for parametrization of atomic force microscopy images (AFM-images). The algorithm provides the formation of a geometric description of AFM-images areas corresponding to the elements of the surfaces of inorganic materials. The stability of the algorithm to changes in the brightness and contrast of the AFM-image is investigated.

Keywords: AFM-images, triangulation, parameterization.

Список литературы

1. Конопелько В.К., Цветков В.Ю. Формирование и обработка образов в помехоустойчивом кодировании и передаче изображений Минск: Бестпринт, 2015.
2. Gwiddion [Электронный ресурс]. URL: http://www.gwiddion.net. (дата обращения: 15.03.2019).
3. Рабцевич В.В., Цветков В.Ю., Ловецкий М.Ю. // Сб. материалов междунар. научн.-техн. конф. «Мониторинг техногенных и природных объектов» Минск: БГУИР, 2017. С. 77–84.

УДК 004.932.72

ПАССИВНАЯ ОПТИЧЕСКАЯ ЛОКАЦИЯ: ВОЗМОЖНОСТИ И ОГРАНИЧЕНИЯ

А.Н. ИСКРИК, А.И. ВИЛЬЧИКОВ, С.В. ГАШКОВ, В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 15 марта 2019

Аннотация. В статье проведен анализ современных методов пассивной локации, применяемых к задаче измерения дальности до наблюдаемого объекта. Описаны основные этапы их работы и особенности реализации. Проведено численное моделирование методов пассивной оценки дальности для оптико-электронных систем.

Ключевые слова: беспилотный летательный аппарат, пассивный метод оценки дальности, оптическая локация.

Введение

На сегодняшний день беспилотные летательные аппараты (БПЛА) являются сложным техническим комплексом позволяющим решать широкий круг задач, таких как построение цифровой карты местности, ведение воздушной разведки в реальном режиме времени, слежение за подвижными объектами, корректировка огня артиллерии, доставка взрывчатых веществ в заданную точку с целью диверсии и многие другие. Малоразмерные воздушные объекты уже несколько десятилетий доставляют множество трудностей системам противовоздушной обороны (ПВО) своими специфическими летно-техническими характеристиками. Низкие значения показателей эффективности поражения малоразмерных БПЛА зенитными средствами обуславливает необходимость разработки и проведения комплекса специальных мероприятий по организации их поражения.

Основные факторы, обусловившие сложность борьбы с БПЛА:

– небольшая масса и габариты и, как следствие, малая дальность обнаружения;

– низкий уровень акустического шума (около 50 дБ на дальностях выше 1000 м, что ниже порога чувствительности органов слуха);

– незначительные величины эффективной площади рассеяния (0,01–0,1 м^2) и тепловой контрастности;

– достаточно широкий диапазон скорости полета (10–30 м/с);

– способность наводить на средства ПВО ударные самолеты, вертолеты и артиллерию;

– возможность полета на предельно малых высотах (до 200 м);

– нечувствительность к психологическому воздействию огня средств ПВО.

Способы определения дальности до цели

Основными способами обнаружения БПЛА являются следующие:

1. Использование тепловизора инфракрасного диапазона.
2. Использование камер оптического диапазона.
3. Использование радиолокационных станций.
4. Осуществление радиомониторинга.

Использование радиолокационных станций для определения дальности до цели

Основным современным средством обнаружения БПЛА являются радиолокационные станции. В ряде случаев БПЛА являются сложной целью для существующих РЛС. Эти аппараты

имеют малую эффективную площадь рассеяния (ЭПР), из-за чего их обнаружение становится достаточно сложной задачей. В частности, снижается максимальная дальность обнаружения [1]. Расчетные дальности обнаружения БПЛА радиолокационными станциями при различных значениях ЭПР БЛА составляют для РЛС метрового диапазона – 8–14 км (БЛА с ЭПР около 0,1 м^2) и 0,1–1,5 км (БЛА с ЭПР, равной 0,01 м^2), для РЛС дециметрового диапазона – 9–16 км (ЭПР = 0,1 м^2) и 0,8–2,0 км (ЭПР = 0,01 м^2); РЛС сантиметрового диапазона – 12–25 км и 1,4–2,8 км [2]. Учитывая установку на них антирадарных покрытий, наибольшую эффективность дает применение двухчастных импульсных радиолокаторов. Первая группа частот в дециметровом диапазоне, вторая – в сантиметровом для обнаружения [3].

Известны следующие способы определения дальности объекта с помощью приемника электромагнитных волн (ПЭВ) [4, 5].

1. Угломерный способ определения дальности объекта и его радиальной скорости с помощью ПЭВ. В этом способе искомые параметры определяются при помощи двух (или более) пассивных ПЭВ, разнесенных на расстояние.

Если объект и два ПЭВ расположены в горизонтальной (вертикальной) плоскости, достаточно найти 2 азимута (2 угла места). Определение дальности осуществляется решением треугольника. Для установления координат объекта в общем случае необходимо измерить не менее 3 угловых координат. Скорость объекта определяется по результатам измерения наклонной дальности в различные моменты времени.

2. Разностно-дальномерный способ определения дальности. Он предусматривает измерение разностей расстояний от излучаемого объекта до ПЭВ. Для установления координат объекта на плоскости требуется определить не менее двух значений разности расстояний, для чего необходимо иметь по крайней мере 3 разнесенных ПЭВ. Местоположение объекта определяется точкой пересечения гипербол, соответствующих измеренным разностям расстояний, с фокусами в точках расположения ПЭВ.

3. Угломерно-разностно-дальномерный способ определения дальности. Он является комбинацией первых двух представленных способов и заключается в определении направлений и разности расстояний от объекта до ПЭВ. При этом способе надо иметь не менее двух ПЭВ. Определение плоскостных координат объекта обеспечивается измерением одной угловой координаты и разности расстояний [6].

Известные способы обладают следующими недостатками:

– не являются однопозиционными;

– для определения дальности и радиальной скорости объекта необходимо применять несколько ПЭВ.

Указанные недостатки приводят к усложнению конструкции измерительной системы и повышению ее стоимости, а также к снижению живучести системы.

Использование оптической локации для определения дальности до цели

Для выявления объектов с отличающейся от окружающей среды температурой используются инфракрасные тепловизионные камеры, что позволяет вести наблюдение за БПЛА в условиях ограниченной видимости и в темное время суток. Для получения наиболее информативных и стабильных результатов возможно точное совмещение тепловизионных снимков с видимым изображением. Для этого применяется тепловизор и видеокамера. Для обнаружения БПЛА в оптическом диапазоне используются активные и пассивные методы. Активными считаются метод анаглифов и метод определения координат БПЛА в пространстве. Пассивные методы включают в свой состав метод визуального наблюдения и метод комбинированного стереоэффекта.

Оптико-электронные системы решают следующие основные задачи.

1. Улучшение видения фоноцелевой картины, выдаваемой телевизионным каналами для повышения эффективности обнаружения и распознавания целей оператором на экране монитора.

2. Отображение на экране монитора фоноцелевой картины от ТВ каналов с наложенной знакографикой.

3. Автоматическое или операторское обнаружение воздушных целей и их захват на автосопровождение.

4. Автосопровождение воздушных целей в подвижном и неподвижном полях зрения.

5. Определение дальности до цели, сопровождение которой ведется в замкнутом контуре слежения.

Общие особенности оптической локации. В оптическом диапазоне облегчается получение координатной информации о цели, ее размерах, форме, ориентации и т. д. При получении координатной информации используют поляризационные и фотометрические характеристики рассеянного излучения, регистрируют изображение цели. Получение координатной информации часто является основной задачей оптических локационных средств. Создание преднамеренных помех для оптической локации возможно, но сложнее, чем для радиолокации.

Полуактивная оптическая локация. Используется явление вторичного излучения (отражения) целями оптических волн от источника естественного интенсивного первичного излучения. Чаще всего таким источником является Солнце. Средства полуактивной локации, основанные на этом принципе, называют оптико-электронными станциями. К средствам полуактивной оптической локации можно отнести также биологические зрительные системы. Пренебрегая фактором использования вторичного излучения, оптико-электронные станции часто относят к средствам пассивной оптической локации.

Пассивная оптическая локация. Использует собственное оптическое излучение нагретых участков поверхности цели или ионизированных образований в ее окрестности. Известно, что максимум излучения абсолютно черного тела при температуре T (по Кельвину) приходится на длину волны ~ 2898/T мкм. Длина волны, на которую приходится максимум излучения реальных целей, обычно находится в инфракрасной области спектра (лишь при $T \sim 4000$ К максимум совпадает с красной, а при $T \sim 5000$ К – с желтой областью видимого спектра). Средства пассивной оптической локации обычно работают в ближнем ИК диапазоне. К подобным средствам относят ИК пеленгаторы, тепловизоры, тепловые головки самонаведения, пассивные приборы ночного видения, видеокамеры и др.

Оптические системы основаны на использовании законов оптики для получения изображений различного масштаба для одного и того же объекта.

Один из алгоритмов определения расположения и размеров объектов заключается в последовательном приближении и удалении экрана или линзы относительно друг друга вдоль оптической оси в некотором заданном диапазоне Δx (рис. 1). В качестве оптической системы может использоваться и более сложная система линз, тогда изображение будет строиться согласно схеме, представленной на рис. 2 [7].

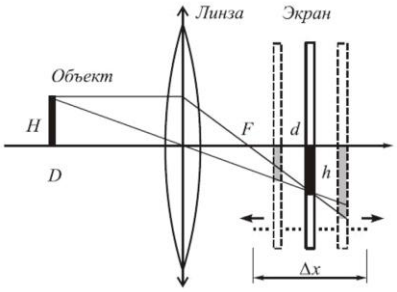


Рис. 1. Формирование изображения на движущемся экране с использованием одной линзы

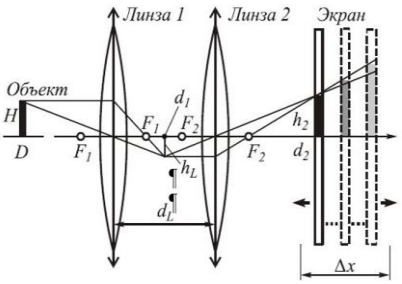


Рис. 2. Формирование изображения на экране с использованием двух линз

В результате расстояние для 1-го случая (рис. 1) определяется по формуле:

$$D_i = \frac{d_i \cdot F}{d_i - F}, \quad (1)$$

где D_i – расстояние от объекта до линзы для *i*-го изображения массива, *F* – фокусное расстояние линзы, d_i – расстояние от линзы до изображения *i*-го кадра массива.

Для 2-го случая значение *D* определяется путем последовательного применения формулы 2 к каждой линзе.

Другой способ заключается в определении расстояния до объекта в дискретные моменты времени посредством оптического фотоприемника, имеющего перестраиваемую оптическую систему с двумя известными граничными фокусными расстояниями (рис. 2) [8]. В результате получаются два изображения различного масштаба и расстояние до объекта определяется по следующей формуле:

$$a = f_2' \frac{1 - \frac{Y_1'}{Y_2'}}{\frac{Y_1'}{Y_2'} \cdot \frac{f_2'}{f_1'} - 1}, \quad (2)$$

где f_1 и f_2 – граничные фокусные расстояния оптической системы; Y_1', Y_2' – размеры изображений объекта при f_1 и f_2 соответственно.

Данные способы имеют высокую точность показания результатов, но наличие механического перемещения в процессе получения изображения влияет на скорость получения результата, как следствие такие системы будут неэффективны при использовании алгоритмов на высоких скоростях передвижения объектов.

Использование двух камер (принцип стереометрии)

Алгоритмы определения расстояния, использующие две камеры, предполагают, что камеры находятся на некотором удалении друг от друга и на одной оси, параллельной относительно изучаемому объекту, а их оптические оси перпендикулярны положения объекта.

Так, например, один из алгоритмов использует две usb-камеры, и, следовательно, стереоскопическое зрение, для построения *Z*-изображения или изображения глубины – одноканального изображения, значение каждого пикселя которого пропорционально дистанции до объекта сцены [9].

Суть метода состоит в следующем. Для каждой точки на одном изображении выполняется поиск парной ей точки на другом изображении. По паре соответствующих точек можно выполнить триангуляцию и определить координаты их прообраза в трехмерном пространстве. На основе трехмерных координат прообраза вычисляется глубина как расстояние до плоскости камеры. Для каждого пикселя левого изображения с координатами (x_0, y_0) выполняется поиск пикселя на правом изображении. При этом предполагается, что пиксель на правом изображении должен иметь координаты $(x_0 - d, y_0)$, где d – величина смещения камер друг относительно друга. В результате получается изображение глубины.

Другой предлагаемый метод измерения расстояния до объекта основан на принципах фотограмметрии и корреляционной обработке цифровых изображений стереопары. Получение информации о дальности заключается в регистрации нескольких изображений объектов под различными ракурсами.

Если расстояние между камерой и наблюдаемой сценой значительно превышает фокусное расстояние оптической системы, можно считать, что изображение строится в ее фокальной плоскости. В проективной модели камеры изображение трехмерного объекта получается его проектированием в фокальную плоскость (рис. 3).

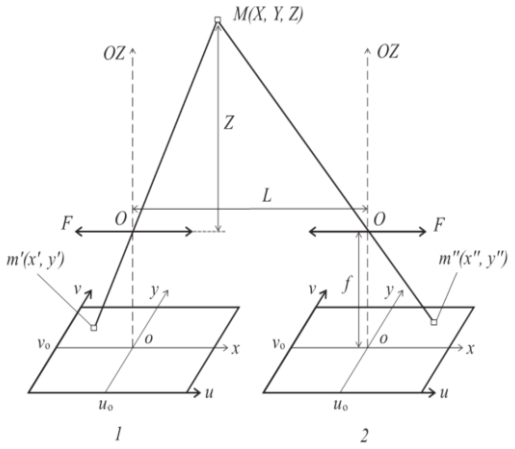


Рис. 3. Проективная модель фотокамеры и стереоскопической системы:
f – фокусное расстояние, L – величина базы, Z – дальность до объекта

Проекцией точки трехмерного пространства M с координатами (X, Y, Z), где Z – расстояние, является точка m с координатами (x, y). При этом выполняются следующие соотношения (3):

$$x = \frac{fX}{Z}; y = \frac{fY}{Z}, \tag{3}$$

где f – фокусное расстояние оптической системы (камеры).

В результате многочисленных вычислений можно установить, что дальность до объекта Z из геометрии изображений определяется соотношением (4):

$$Z = \frac{(f \cdot L)}{(x' - x'')}, \tag{4}$$

где f – фокусное расстояние, L – величина базы (расстояние между оптическими осями камер), x' и x'' – координаты объекта в плоскости изображения на правом и левом снимках стереопары соответственно, а их разность – смещение объекта в плоскости изображения на первом и втором снимках стереопары соответственно.

Для более удобного практического применения формулы (4) представим ее в виде

$$r_н = \frac{LH}{tg\alpha(x_1 - x_2)}, \tag{5}$$

где L – величина базы (расстояние между оптическими осями камер); H – горизонтальное разрешение изображения; α – угол обзора камеры; x_1 и x_2 – координаты точки, до которой определяется расстояние, в координатной системе отсчета первой и второй камеры соответственно [10].

Для возможности использования формулы (5) считается, что изображения, получаемые с камер, ректифицированы, т. е. камеры расположены так, что в их координатных системах отсчета координаты точки, до которой требуется определить расстояние, равны. Это означает, что горизонтальные линии на изображениях соответствуют одной плоскости.

Из условия, в соответствии с которым чувствительность определения расстояния должна быть высокой (т. е., чтобы изменение разности пикселей на единицу приводило к изменению определяемого расстояния не более чем на 5 %), можно определить, начиная с какой разности пикселей следует применять формулу (4):

$$r_1 - r_2 = 0,05r_1,$$

$$0,95 \cdot L \cdot H / (\text{tg}a \cdot Dx_1) = L \cdot H / (\text{tg}a \cdot Dx_2),$$

$$\Delta x_1 = 0,95 \cdot \Delta x_2.$$

Учитывая, что $\Delta x_2 = \Delta x_2 + 1$, получаем $\Delta x_1 = 19$.

Это означает, что, начиная с разности пикселей двух изображений, равной 19, возможно применять метод определения расстояния до объекта с помощью стереовидения.

Для того чтобы можно было определить расстояние вплоть до 500 м при угле обзора одной камеры $\alpha = 1{,}5°$ и горизонтальном разрешении изображения, равном 1920 пикселей, вычислим по формуле базу *L*:

$L = r_и \cdot \mathrm{tg}\alpha \left(x_1 - x_2 \right) / H = 1{,}14$ м.

На рис. 4 показана зависимость расстояния до исследуемого объекта от разности между изображениями, полученными с двух стереокамер.

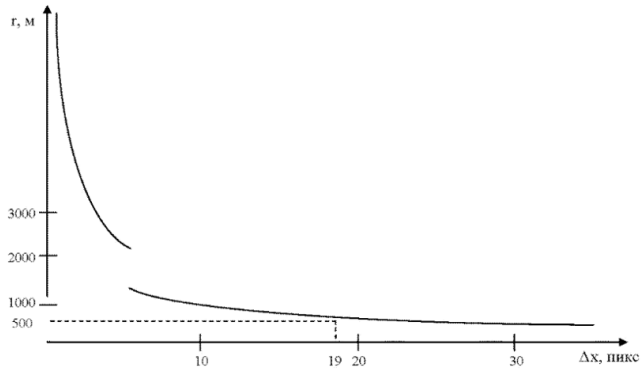


Рис. 4. Зависимость расстояния до объекта от разности пикселей между изображениями объекта с двух камер

PASSIVE OPTICAL LOCATION: FEATURES AND LIMITATIONS

A.N. ISKRYK, A.I. VILCHIKOV, S.V. GASHKOV, V.Yu. TSVIATKOU

Abstract. The current methods of passive location are analyzed. The main stages of their work and features of implementation are described. A numerical simulation of passive range estimation methods for optoelectronic systems is carried out.

Keywords: unmanned aerial vehicle, passive range estimation methods, optical location.

Список литературы

1. Рябов К. Как противодействовать беспилотнику // Военное обозрение. [Электронный ресурс] URL: https://mensby.com/technology/guns/5386-how-counteract-drone (дата обращения: 15.03.2019).
2. Годунов А.И., Шишков С.В., Юрков Н.К. // Надежность и качество сложных систем. 2014. № 2 (6). С. 62–70.
3. Свинин Е.В. Рамочная антенна для системы радиомониторинга. Выпускная квалификационная работа специалиста. Санкт-Петербург. 2017.
4. Быстров Р.П., Загорин Г.К., Федорова Л.В. Пассивная радиолокация: методы обнаружения объектов. М.: Радиотехника, 2008.
5. Охрименко А.Е. Основы радиолокации и радиоэлектронная борьба. Ч.1. Основы радиолокации. М.: Воениздат. 1983. 456 с.
6. Черевко А.А., Ильин Е.М., Черевко А.Г, Способ пассивного однопозиционного определения дальности объекта и его радиальной скорости. 2018, [Электронный ресурс] URL: https://edrid.ru/rid/218.016.1346.html (дата обращения: 15.03.2019).
7. Гейдаров П.Ш. // Компьютерная оптика. 2011. Том 35. № 2. С. 275–280.
8. Моностатический способ определения расстояния до объекта, его направления и скорости движения: пат. 2340872 РФ/ Зуев С.В.; опубл.: 10.12.2008.
9. Козлов В.Л., Кузьмичев И.Р. // Вестник БГУ. 2011. Сер. 1. № 1. С. 33–38.
10. Локтев Д.А., Алфимцев А.Н. // Инженерный журнал: наука и инновации. 2013. Вып. 11. URL: http://engjournal.ru/catalog/ it/hidden/996.html (дата обращения: 15.03.2019).

УДК 004.932.2

ОБНАРУЖЕНИЕ И ДЕТЕКТИРОВАНИЕ ОБЪЕКТОВ НА ВИДЕОИЗОБРАЖЕНИИ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Д.Ю. ТУЧА

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 13 марта 2019

Аннотация. Предложено программное обеспечение, разработанное на языке Python 3 для обнаружения и детектирования объектов на кадрах видеоизображения на основе алгоритма SSD (от англ. Single Short Derector) и библиотеки TensorFlow. Проведена практическая оценка предложенного программного обеспечения.

Ключевые слова: обнаружение объекта, видеоизображение, машинное обучение, нейронная сеть.

Введение

В последнее время широкое распространение робототехники, систем автоматического управления, видеонаблюдения, дополнительной реальности и других систем, связанных с распознаванием объектов на изображении, вызвало обширные исследования в области компьютерного зрения. Благодаря значительному развитию нейронных сетей, особенно глубокого обучения, системы визуального распознавания достигли значительных результатов. В работе предложена программная реализация метода обнаружения объектов с помощью глубокого обучения на основе языка программирования Python 3.

Машинное обучение

Программа реализована на основе открытой программной библиотеки для машинного обучения TensorFlow. В библиотеке вычисления описываются с использованием графов потока данных (рис. 1).

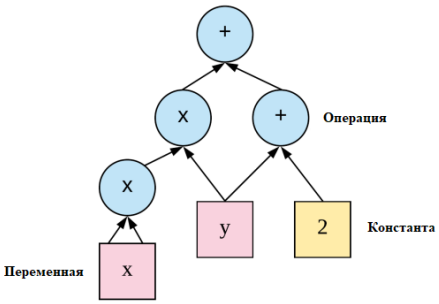


Рис. 1. Граф потока данных

Каждый узел графа представляет собой математическую операцию (например, сложение, деление или умножение), а каждое ребро – многомерный набор данных (тензор), над которым выполняются операции [1]. Тензор представляет собой многомерную матрицу, заполненную числами – компонентами тензора.

Модель глубокого обучения для обнаружения объектов можно описать тремя шагами:

– сбор данных для обучения;

– обучение модели;

– предсказание на новых изображениях.

Применение метода Single Shot Detector

Для обнаружения объектов на видеоизображении используется алгоритм SSD. Обнаружение состоит из извлечения карты признаков и применения свертки для обнаружения объектов. На вход подается изображение размером 300×300 пикселей. К изображению применяются сверточные слои из стандартной модели VGG-16. Далее к каждому выходному слою добавляются специальные сверточные слои, представляющие собой изображение в разных масштабах. Пространственная размерность убывает до тех пор, пока не станет равной единице. Каждый из сверточных слоев позволяет сформировать карту признаков для разных масштабов изображения, например 8×8 (рис. 2) и 4×4 (рис. 3). Далее для карты признаков определяется, какой из ограничивающих прямоугольников в области 3×3 лучше всего совпадает с эталонной разметкой. Все эти карты объединяются в единый выходной слой, в которых может находиться объект. Итоговые области выбираются из прямоугольников с помощью метода подавления немаксимумов [2].

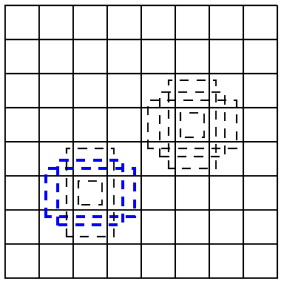


Рис. 2. Карта признаков для масштаба изображения 8×8

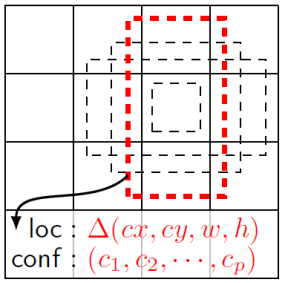


Рис. 3. Карта признаков для масштаба изображения 4×4

Оценка эффективности разработанной программы

Качество обнаружения транспортных средств описывают такие метрики, как полнота R и точность P обнаружения объектов [3]. Полнота R показывает чувствительность алгоритма к ошибкам 2-го рода, то есть, пропускам, и равна отношению количества правильно найденных объектов к общему количеству этих объектов в эталонной разметке:

$$R = \frac{tp}{tp + fn},$$

где tp – истинно-положительные – объекты, которые ожидалось увидеть и которые были получены на выходе, fn – ложноотрицательные объекты, которые ожидалось увидеть, но алгоритм их не определил.

Точность P показывает чувствительность алгоритма к ошибкам 1-го рода, то есть ложным срабатываниям, и равна отношению количества правильно найденных объектов к общему количеству найденных алгоритмом прямоугольников:

$$P = \frac{tp}{tp + fp},$$

где fp – ложноположительные объекты, которых на выходе не должно быть, но алгоритм их ошибочно вернул на выходе.

Оценка эффективности разработанной программы проводилась на кадрах изображений, полученных с автомобильного видеорегистратора (рис. 4).

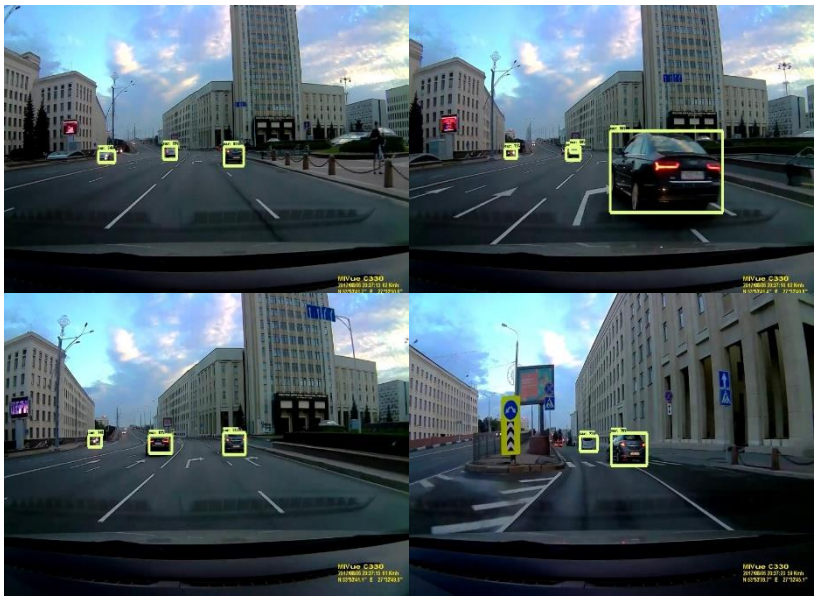


Рис. 4. Примеры обнаружения транспортных средств

В первом тесте использовалась уже обученная нейронная сеть на выборке ssd_mobilenet_v1_coco. Второе тестирование проводилось на дополнительно обученной выборке из первого теста. Полученные результаты представлены в таблице.

Результаты обнаружения транспортных средств на готовой выборке и обученной выборке

Номер теста	Полнота	Точность
Тест 1	0,464	0,890
Тест 2	0,993	0,989

Заключение

Разработано программное обеспечение на языке Python 3, которое позволяет обнаружить и детектировать объекты на видеоизображении. На основе результатов его испытаний установлено, что чем больше различных изображений объектов используется при обучении нейронной сети, тем выше результаты качества распознавания. Имеется возможность улучшить производительность работы программного обеспечения путем перехода на графические процессоры с использованием технологии nVidia CUDA.

OBJECTS DETECTION AND RECOGNITION ON VIDEOIMAGES USING MACHINE LEARNING

D.Yu. TUCHA

Abstract. Software for object detection and recognition on video images using algorithm SSD (Single Shot Detector) and open source software library TensorFlow is proposed.

Keywords: detection, videoimage, machine learning, neural network.

Список литературы

1. Ramsundar B., Zadeh R.B. // O' Reilly Media Publishers. 2018. Vol. 2. P. 19–42.
2. SSD: Single Shot MultiBox Detector. [Электронный ресурс]. URL: https://arxiv.org (дата обращения: 15.03.2019).
3. Everingham M. [et. al.] // Int. J. of Computer Vision. 2010. Vol. 88. P. 303–338.

СЕКЦИЯ 3
ЗАЩИТА ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И СЕТЕЙ

УДК 007.681.512.2

TECHNOLOGY OF ADAPTIVE INTELLIGENCE MULTIAGENT PROCESSING FOR INFORMATION DEFENSE SYSTEM DESIGN

U.A. VISHNIAKOU, A.H. Al-MASRI, S.K. Al-HAJI

Belarusian state university of informatics and radioelectronics, Republic of Belarus

Submitted 18 March 2019

Abstract. The tendencies of using multi-agent intelligent technologies for information processing are given. The main ideas of building a distributed multi-agent system with distributed knowledge and distributed processing are shown. The directions of multi-agent intelligent systems utilizationin information defense using cloud technologies are presented. Knowledge representation network in the form of object models is proposed, on basic which the intelligent MAS should be constructed. The approaches for adaptive intelligent multi-agent technology in area of information security are introduced.

Keywords: multi-agent technologies, distributed knowledge bases, distributed decision-making, cloud environment.

Introduction

One of the main problems of innovative economy building is intellectualization, the essence of which is to develop effective mechanisms for the formation, publication, updating and mass use of innovative knowledge in management technologies. Among such knowledge are allocated developments in the field of intellectual agents on the basis of semantic Web, Web services and semantic Web services, cloud computing, blockchain technologies [1].

The technology of development and use of multi-agent systems (MAS) and multi-agent management (MAM) is understood like under multi-agent technology. The problems of control and distributed interaction in networks of dynamic systems attract the attention of a large number of researchers. This is due to the widespread use of multi-agent systems in different areas, including automatic adjustment of parameters of neural recognition networks, transport management, distributed sensor networks, control in communication networks, interaction of UAV groups, management of mobile robots, protection of information resources, etc. Distributed MAS are used that perform actions in parallel, for which the task is dividing on parts between several computational threads. Such problems arise not only in computer networks, but also in production networks, service networks, transport and logistics networks. With natural constraints on communication, decentralized strategies are able to effectively solve this type of problem [2, 3].

The basic of MAC

Multi-agent systems originated at the intersection of system theory and distributed artificial intelligence. Open, active, developing systems are discussed, in which attention is paid to the processes of agents interaction for building systems with new qualities. MAS are built as a union of individual intelligent systems based on knowledge [3]. MAS consists of the following components: a set of agents working with objects; variety of tasks; a space in which there are agents and objects; the set of relations

between agents; many agent actions (operations on objects). Agent management system (AMS) is also an agent that controls access and use of the agent platform [2].

The basis of the organization form of interaction between agents characterized by the combination of their efforts to achieve the goal in the division between their functions, roles and responsibilities is cooperation (C). This can be determined:

K = cooperation + coordination + communication.

It's means the management of the associations between actions under the coordination. Communication between agents depends on the chosen protocol, which is a set of rules that determine how to synthesize meaningful and correct messages.

In the MAC architecture, the main part is the domain-independent core, which includes such components: direct access service (provides direct access to the attributes of agents); message service is responsible for the transmission of messages between agents and kernel systems; agent class library (part of the database) contains the classification of agents in the MAS; agents community, where agents are located (this block provides functions for loading/writing agents and their properties and optimizes the work of agents with resources); ontology is a subject knowledge base containing specific knowledge about objects and environment of functioning, represented in the form of a corresponding semantic network [2].

The agent structure and their use

The basis of agent structure is the context, or server environment, in which it is executed. Each agent has a fixed identifier-name. In a server environment, you can run not only the source agent, but also a copy of it. Agents are able to create their own copies, sending them to different servers for execution. When the agent arrives on the next server, its code and data are transferred to the new context and erased at the previous location. In the new context the agent can do anything that is not prohibited there. Upon completion of the work in the context the agent may send itself in a different context or upload sender address. Agents can also shut down themselves or at the command of the server, which then moves them from the context to the storage location.

The structure of a typical agent includes inputs (internal parameters of the agent and data on the state of the environment), outputs (parameters affecting the environment and informing the user about the state of the environment and decisions made), the solver – the decision-making procedure. The solver can be a fairly simple algorithm or an element of an artificial intelligence system [3].

Multi-agent systems are used for the development of information and industrial systems. In industry the MAS are used to the solution of management automation of complex systems, for the collection and processing of information in games. Multi-agent technologies are applicable in the management of mobile resources, as well as in such areas as object design, industrial production [2].

In sources [4, 5] MAS deals with the application of automate the construction of intelligent knowledge bases and problem solvers. The resulting model of hybrid knowledge bases, which ensures the compatibility of present knowledge and can be represented in knowledge bases multi-level meta knowledge, to structure the knowledge base according to various criteria and to apply components of knowledge bases again [4]. The agent-oriented model of the hybrid solver allows to build variety of MAS: for production, customer service, construction design [5].

Design of MAS

The general methodology of the ascending evolutionary design of MAC can be represented by a chain: <environment – functions OF Mac – role of agents – relations between agents – basic structures of MAC-modification>. It includes the stages of: formulation of purpose (objectives of development) MAC; the identification of core and support functions of agents; clarify the composition of agents and the distribution of tasks among agents, the choice of the architecture of the agents; the provision of basic relationships between agents; determination of possible actions (operations) agents; analysis of real-life, real or anticipated changes in the environment. When designing, the organization of agents can be considered as a set of roles that are in a certain relationship with each other and interact with each other [2].

MAS bottom-up design methodology requires a preliminary task of the initial functions, determining the range of their obligations to each other, the formation of the initial structures and its developing on the basis of the allocated functions and the study of the adequacy of these structures to the nature of the tasks in the selected problem areas.

The technique of top-down design is to determine the social characteristics of MAS on a set of criteria, the construction of the basic types of their organizations, followed by the definition of requirements for the architecture of agents. For artificial social systems and communities, a top-down approach to organizational design is put forward [3].

Agents can be integrated into cloud computing (CC) structures that contain specific functions for problem solving, data processing, and management. They support a natural mix of knowledge-based information and technology and can support the process of logical reasoning (for example, including business regulations). They enable learning and self-improvement at both the infrastructure level (adaptive routing) and the application level (adaptive user interfaces) [6, 7].

There are several international approaches to creating a MAS, the most famous of them are [2]: MASIF (Object Management Group), which is based on the concept of a mobile agent; FIPA (Foundations for Intelligent Physical Agents) specifications based on the intelligence of the agent, as well as standards developed by the research subsection Defense Advanced Research Projects Agency (DARPA), in particular Control of Agent Based Systems. Regarding mobility and intelligence of agents, most experts agree that mobility is the Central characteristic of the agent, intelligence is desirable, but not always strictly required [2, 3].

FIPA's activities include joint research and development by its members of international specifications that will maximize the interaction between agent applications, services and equipment. FIPA specifications focus on enabling intelligent agent communication through standardized agent communication and content languages. Along with the General basics of communication, FIPA also specializes in ontology and negotiation protocols to support interaction in specific application areas (transport support, production, multimedia, networking) [2].

The OMG MASIF standard creates conditions for the migration of mobile agents between MAS via standardized CORBA IDL interfaces. DARPA initiated the work on the distribution of Knowledge Sharing Effort, as a result of which the agent programming languages were divided into syntax, semantics and pragmatics: language KIF (Knowledge Interchange Format) – syntax; Ontolingua – language for defining shared ontology's (semantics); KQML (Knowledge Query and Manipulation Language) – a high-level interaction language (pragmatics). When you create a MAS is also used language of communication between agents – Agent Communication Language (ACL) that specifies the types of messages agents, the content and the ontology. Cooperation between agents is achieved through a set of basic concepts used in communications. The ontology is used as the Application Programming Interface and defines the agent interface.

Within the framework of design MAC direction, architectures, models and software prototypes of several MAS were developed: attack modeling, intrusion detection, intrusion detection training, etc. [8].

Knowledge representation network in the form of object models

Introduction of concepts of objects-carriers and objects-communicators, structure, operations over them, the recursive description and levels of abstraction allows to define the abstract machine of objects on the basis of which it is possible to create this or that object computing environment with software or hardware implementation. This will create effective implementations for information management and security [7]. Usually the knowledge base (KB) is a set of facts and rules that can be changed during processing. Consider the KB represented by a set of special *n*-objects [9].

By definition, an *n*-object is a pair of $<A_i, f_{i1}>$ two *n*-components. Next, we will use only those objects in which $A = \{0, 1\}$ and f_{ij} is a predicate. Each predicate defines a set of objects such as P_n ($n \geq 2$). Each object in the P_n set has an identifier U denoting its class (class) and name (name), i. e. $U <A_1, \ldots A_n>$ and $<f_{i1}, \ldots f_{in}>$.

Each f_{ji} function from an *n*-object will be treated as a pair of $f_{ij}\ f_{ij}$*, where f_{ij} is an access function for a set of A_i elements, and f_{ij}* is a function that assigns a value to the argument xi in the next step of the operation. Suppose that all f_{ij} functions (predicates) in a given n-object depend on one set of arguments $\{x_1, x_2 \ldots x_m\}$. This assumption allows us to call the considered n-object set P_n,

with the above properties, the knowledge base, if for any value of the argument only one predicate from the set of predicates is true for each set of argument values. Processing of the object knowledge base (OKB) is performed in the following way:

1. It must set the original values of the x_1 argument ... x_n. This locks the object in which f_{i1} (x_1, x_2... x_n) = 1. At this step, facts from set A_2 ... A_n can be used.

2. For each object all values of f_{ij} functions (predicates) are calculated (for all $j \geq 2$), which generate values x_1, x_2... x_n in the next step.

The object-logical model of knowledge representation has two levels:

– the first level includes an object-oriented language of knowledge description;

– the second level is a system that includes an object base (OKB), an object network and a logical subsystem. OKB has descriptions of all classes of objects. An object-logical network combines the capabilities of functional and semantic networks, and can include objects and relationships for some application area.

Classes, name, value, or set of values are connected to each network object. Relations / functions calculate or correct the values of their arguments and try to join object nodes, otherwise object nodes, provided that the arguments are identical.

Such local processes will be initiated (in other words, will start, will run) at any change of object nodes and will be executed until the complete stabilization of the network is achieved. Many rules are connected to any node of the object network. These rules are triggered when the object node is modified.

Problem solving with object networks

The following problems can be formulated and solved on the basis of the presented object-logical network, using knowledge from monograph [9]:

1. Design objects with the required properties, characteristics, constraints, and so on, using the original objects from the BR;

2. Test the object and its components to find possible failures in areas such as computers, telecommunication system and etc.;

3. To control complex objects;

4. To simulate a certain action in the technical and social systems;

5. To carry out examination of some problems and to make a decision in the conditions of incomplete or contradictory information.

Addressing these issues will include the following steps:

1. Identifying the problem when using the high-level object logic language (HOLL);

2. Translation of the applied description of the problem into an internal form and creation (on its basis) of a working network of objects using OKB.

3. Transform the resulting object-logical network by calculating the appropriate algorithm.

Adaptive intelligent MAC concept

The report presents the following solutions for adaptive intelligent multi-agent information security technology (AI MAT ID):

− the subsystem of analysis of information security of the protected object, implementing the audit procedures for the state of protection of the object;

− attack detection system, which includes agents of workstations, servers, routers and subnets and allows to draw a conclusion about the presence and variety of attacks;

− dynamic knowledge base of is agents, including ontologies, formal and heuristic rules, BS interface;

− the method of making a joint decision by agents, which allows to assess the state of information security of the protected object on the basis of a dynamic knowledge base formed from various sources;

− multi-agent attack counter subsystem, which allows to accumulate knowledge and train a multi-agent system for further detection of new threats;

− subsystem for evaluating the effectiveness of the proposed methods of protection.

The architecture of MAS is built using this technology and includes a knowledge base in the form of rules of production, the mechanism of inference, receptors and effectors of the agent, the module of communication with other agents. In relation to the task of security analysis, agents transmit facts about external influences to the knowledge base. As a result of the logical output, a solution is developed, which is transmitted to the processor about changes in the external environment. Different types of agents can be used for distributed problem solving: agent-subordinator, multiple agents of executors, agent-integrator. Agents can be interconnected as a multi-level architecture, which can be horizontal or vertical.

Conclusion

1. The direction of intelligent MAS construction the further development of models, methods, architecture and software to solve the problem of adaptive information security system design.

2. Knowledge representation network in the form of object models is proposed, on basic which the intelligent MAS shell be construct.

3. The approaches for adaptive intelligent multi-agent technology (AI MAT) in information security area are introduced.

References

1. Vishniakou U.A. Information management and security: methods, models, software and hardware solutions. Monograph. Minsk. MIU, 2014.
2. Leyton-Brown K, Shoham. Y Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations London: Cambridge University Press. 2009.
3. Rzevski G. // Proc. of 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD2012). Kyoto, Japan, August 8–10. 2012. P. 434–437.
4. Davydenko I.T. Models, methods and means of development of generated knowledge bases on the basis of semantic compatibility of reusable components: autoref. of PhD thesis in tech. sciences Minsk, 2018.
5. Shunkevich D.V. Agent-oriented problem solvers of intelligent systems component: autoref. of PhD thesis in tech. sciences Minsk, 2018.
6. Vishniakou U.A. Information security in corporate systems, electronic commerce and cloud computing: methods, models, hard-software tools Minsk, Bestprint, 2016.
7. Fingar P. Cloud computing – the business platform of the XXI century M.: Aquamarine Book, 2011.
8. Kotenko I.V., Yusupov R.M. // Bulletin of the Russian Academy of Sciences. 2007. Vol. 77. No. 4. P. 323–333.
9. Vishnyakou U. A., German O.V. Models and Tools of Logical Inference Systems. Moscow: Nauka, 1999.

УДК 004.056.53

ОЦЕНКА ПОТЕНЦИАЛЬНЫХ УЯЗВИМОСТЕЙ ИНФОРМАЦИОННЫХ СИСТЕМ И СЕТЕЙ ОРГАНИЗАЦИЙ ЭЛЕКТРОСВЯЗИ

В.А. БОЙПРАВ, Л.Л. УТИН

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 марта 2019

Аннотация. Проанализированы потенциальные уязвимости информационных систем и сетей, используемых организациями электросвязи. В соответствии со стандартом CVSS v3 выполнена оценка этих уязвимостей. Представлен анализ этой оценки.

Ключевые слова: организации электросвязи, угроза, уязвимость.

Введение

Одним из наиболее важных критериев, используемых для оценки защищенности информационных систем и сетей, является качество системы ее мониторинга, которая представляет собой совокупность мероприятий, направленных на выявление потенциальных угроз информационной безопасности. Важность указанного критерия обусловлена тем, что на нем базируется процесс построения эффективной системы защиты информации, который включает в себя определение комплекса организационных и технических мер, направленных на обеспечение конфиденциальности, доступности, целостности, подлинности и сохранности данных, циркулирующих в информационной системе или сети, а также принципы и порядок (очередность) внедрения этих мер.

Очередность внедрения каждой из мер как элемента системы защиты информации определяется критичностью реализации потенциальной угрозы информационной безопасности, на устранение которой направлена эта мера. Для определения указанной очередности необходимо в процессе мониторинга информационных систем и сетей выполнять построение моделей потенциальных угроз безопасности циркулирующих в них данных. Оно заключается в описании следующих свойств каждой из таких угроз [1]: уязвимость информационной системы и сети, на которую воздействует угроза; источник угрозы; способ реализации угрозы (принцип воздействия угрозы на уязвимость); последствия от реализации угрозы.

Первое из указанных свойств является базовым для определения критичности реализации потенциальной угрозы информационной безопасности.

В рамках представляемой работы выполнена оценка уязвимостей информационных систем и сетей организаций электросвязи, для чего были решены следующие задачи:

1. определение перечня потенциальных уязвимостей информационных систем и сетей организаций электросвязи;
2. выбор и обоснование методики оценки потенциальных уязвимостей;
3. анализ полученных результатов оценки.

Актуальность представляемой работы обусловлена тем, что информационные системы и сети организаций электросвязи являются критически важными объектами информатизации (КВОИ) либо элементами инфраструктуры таких объектов, в связи с чем от степени защищенности этих систем и сетей зависит состояние национальной безопасности.

Анализ потенциальных уязвимостей информационных систем и сетей организаций электросвязи

В работе [2] выделены следующие группы активов информационных систем и сетей организаций электросвязи: аппаратные, программные, аппаратно-программные, обрабатываемая

информация, информационные процессы. В табл. 1 представлен перечень основных активов информационных систем и сетей организаций электросвязи различных сфер деятельности.

Таблица 1. **Активы информационных систем и сетей организаций электросвязи различных сфер деятельности**

Сфера деятельности организации электросвязи	Наименование группы активов информационных систем и сетей	Основные активы информационных систем и сетей
Проектирование сетей и сооружений электросвязи	Программные	Программное обеспечение, используемое для проектирования сетей и сооружений электросвязи
	Аппаратно-программные	Средства вычислительной техники
	Обрабатываемая информация	Схемы разработанных или разрабатываемых проектов
	Информационные процессы	Создание, хранение, передача
Строительство сетей и сооружений электросвязи	Аппаратные	Линии и средства электросвязи
Предоставление услуг электросвязи	Аппаратные	Средства электросвязи
	Программные	Программное обеспечение для работы с базами данных и для защиты информации
	Аппаратно-программные	Средства вычислительной техники
	Обрабатываемая информация	Данные об абонентах
	Информационные процессы	Создание, хранение
Государственное регулирование и управление в области электросвязи	Программные	Программное обеспечение для организации электронного документооборота и защиты информации
	Аппаратно-программные	Средства вычислительной техники
	Обрабатываемая информация	Документы, регулирующие деятельность в области электросвязи
	Информационные процессы	Создание, хранение, передача

Потенциальные уязвимости информационных систем и сетей обусловлены уязвимостями их активов, а также несоблюдением пользователями этих систем и сетей организаций требований о неразглашении сведений, связанных с особенностями их функционирования и эксплуатации, а также с содержанием обрабатываемой информации.

Потенциальные уязвимости аппаратных и аппаратно-программных активов могут быть связаны со следующими особенностями их реализации и эксплуатации:

– выход из строя ввиду производственных дефектов;

– некорректное подключение;

– несвоевременность устранения повреждений;

– незащищенность от побочного электромагнитного излучения.

Причинами потенциальных уязвимостей программных активов могут быть следующие:

– ошибки в программных кодах, позволяющие злоумышленнику реализовать несанкционированный доступ в информационную систему или сеть либо внедрить вредоносное программное обеспечение;

– ошибки в программных кодах, приводящие к сбою процессов функционирования программных активов;

– некорректность настроек;

– несовместимость программных активов друг с другом (в том числе с используемыми для управления информационными системами и сетями операционными системами).

Потенциальные уязвимости информационных процессов связаны как с представленными выше уязвимостями, так и с несовершенством используемых алгоритмов шифрования, ошибками проектирования информационной системы или сети.

Методика выполнения оценки потенциальных уязвимостей

Для выполнения оценки потенциальных уязвимостей информационных систем и сетей организаций электросвязи использован стандарт CVSS v3 (от англ. Common Vulnerability Scoring

System version 3). Выбор указанного стандарта обусловлен большим количеством предусмотренных в рамках него метрик уязвимости и соответствующих им характеристик, что позволяет обеспечить высокую точность ее оценки.

Метрики, предусмотренные в стандартах CVSS, делятся на следующие виды:

– базовые (совокупность характеристик уязвимости, не меняющихся со временем);

– временные (совокупность характеристик уязвимости, используемых для описания полноты имеющейся о ней информации, степени зрелости эксплуатирующего ее программного кода);

– контекстные (совокупность характеристик информационной системы или сети, в которой обнаружена уязвимость).

В табл. 2 представлены характеристики, соответствующие метрикам указанных видов, а также описание этих характеристик.

Таблица 2. **Характеристики метрик, предусмотренных стандартом CVSS v3**

Наименование метрик в зависимости от их вида	Наименование характеристики	Обозначение характеристики	Параметры характеристики
Базовые	Attack vector (вектор атаки)	AV	Network (N) Adjacent Network (A) Local (L) Physical (P)
	Attack complexity (сложность атаки)	AC	Low (L) High (H)
	Privileges required (требуемый уровень привилегий)	PR	High (H) Low (L) None (N)
	User interaction (необходимость взаимодействия с пользователем)	UI	None (N) Required (R)
	Scope (границы эксплуатации)	S	Unchanged (U) Changed (C)
	Confidentiality impact, integrity impact, availability impact (метрики воздействия)	C I A	None (N) Medium (M) High (H)
Временные	Exploit code maturity (степень зрелости доступных средств эксплуатации уязвимости)	E	Not Defined (ND/X) High (H) Functional (F)[1] Proof-of-Concept (POC/P)[2] Unproven (U)[3]
	Remediation level (доступные средства устранения уязвимости)	RL	Not Defined (ND/X) Unavailable (U) Workaround (W)[4] Temporary Fix (TF/T)[5] Official Fix (OF/O)[6]
	Report confidence (степень доверия к информации об уязвимости)	RC	Not Defined (X) Unknown (U)[7] Reasonable (R)[8] Confirmed (C)[9]
Контекстные	Confidentiality requirement, integrity requirement, availability requirement (требования к безопасности)	CR IR AR	Not Defined (ND/X) High (H) Medium (M) Low (L)

[1] Имеется программный код для эксплуатации уязвимости
[2] Имеется сценарий реализации атаки
[3] Наличие программного кода для эксплуатации уязвимости не подтверждено
[4] Средства устранения уязвимости разработаны самой организацией (являются неофициальными)
[5] Средства устранения уязвимости являются временно официальными
[6] Средства устранения уязвимости являются официальными
[7] Описание причины уязвимости отсутствует
[8] Существуют отчеты об уязвимости, с помощью которых можно установить причины ее возникновения, а также выполнить ее оценку
[9] Наличие уязвимости подтверждено производителем продукта

Результаты и их обсуждение. Выводы

В табл. 3 представлены вектора наиболее критичных потенциальных уязвимостей информационных систем и сетей организаций электросвязи различных сфер деятельности.

Таблица 3. **Вектора потенциальных уязвимостей информационных систем и сетей организаций электросвязи различных сфер деятельности**

Сфера деятельности организации электросвязи	Вектора основных потенциальных уязвимостей информационных систем и сетей		
	Базовые метрики	Временные метрики	Контекстные метрики
Проектирование сетей и сооружений электросвязи, предоставление услуг электросвязи, государственное регулирование и управление в области электросвязи	AV:N/AC:H/PR:L/UI:R/S:C/C:H/I:H/	E:F/RL:O/RC:R	CR:H/IR:H/AR:H
Строительство сетей и сооружений электросвязи	AV:P/AC:L/PR:L/UI:N/S:C/C:N/I:H/A:H	E:F/RL:W/RC:X	CR:H/IR:M/AR:M

Информационные системы и сети организаций электросвязи, сфера деятельности которых сопряжена с проектированием, предоставлением услуг и государственным регулированием, характеризуются одинаковыми векторами наиболее критичных потенциальных уязвимостей, которые обусловлены уязвимостями программных активов.

На основании анализа полученных векторов потенциальных уязвимостей информационных систем и сетей организаций электросвязи бала выполнена их оценка [3]. В табл. 4 представлены результаты этой оценки.

Таблица 4. **Оценки потенциальных уязвимостей информационных систем и сетей организаций электросвязи различных сфер деятельности**

Сфера деятельности организации электросвязи	Оценки основных потенциальных уязвимостей информационных систем и сетей		
	Базовые метрики	Временные метрики	Контекстные метрики
Проектирование сетей и сооружений электросвязи, предоставление услуг электросвязи, государственное регулирование и управление в области электросвязи	8,0	7,1	7,6
Строительство сетей и сооружений электросвязи	7,1	6,7	6,7

Из табл. 4 следует, что в соответствии со стандартом CVSS v 3, потенциальные уязвимости информационных систем и сетей организаций электросвязи, сфера деятельности которых сопряжена с проектированием, предоставлением услуг и государственным регулированием являются более критичными, чем потенциальные уязвимости систем и сетей, эксплуатируемых организациями, задействованными в строительстве сетей и сооружений электросвязи. Однако класс вторых из упомянутых информационных систем и сетей, как КВОИ, выше, чем класс первых. В связи с этим, по мнению авторов, для того, чтобы использовать стандарт CVSS v 3 для оценки потенциальных уязвимостей информационных систем и сетей организаций электросвязи необходимо использовать поправочные коэффициенты, зависящие от класса этих систем и сетей как КВОИ. Дальнейшие исследования будут направлены на определение и обоснование этих коэффициентов.

ASSESSMENT OF POTENTIAL VULNERABILITIES OF TELECOMMUNICATION ORGANIZATIONS' INFORMATION SYSTEMS AND NETWORKS

V.A. BOIPRAV, L.L. UTIN

Abstract. The potential vulnerabilities of information systems and networks used by telecommunication organizations is analyzed. These vulnerabilities were assessed in accordance with the CVSS v3 standard. An analysis of this assessment is presented.

Keywords: telecommunication organizations, threat, vulnerability.

Список литературы

1. ТКП 483-2013 (01019). Информационные технологии и безопасность. Безопасная эксплуатация и надежное функционирование критически важных объектов информатизации. Общие требования. Минск: ОАЦ, 2013. 6 с.
2. Бойправ В.А., Утин Л.Л. Особенности подготовительного этапа аудита системы менеджмента защиты информации в организациях электросвязи // Докл. БГУИР. 2018. № 1 (111). С. 43–50.
3. NVD – CVSS v3 Calculator [Electronic resource]. URL: https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator (date of access:15.03.2019).

УДК 537.531:621.039.537-037.87

КОМПОЗИТ ДИОКСИДА ТИТАНА И УГЛЕРОДА ДЛЯ СОЗДАНИЯ ЭКРАНОВ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

М.С.Х. Аль-МАХДАВИ, Е.С. БЕЛОУСОВА

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 11 марта 2019

Аннотация. Представлены исследования процессов внедрения частиц технического углерода и диоксида титана в состав синтетического материала. Показано изменение частотных характеристик коэффициентов передачи и отражения, измеренных в режимах согласованной нагрузки и короткого замыкания, в результате введения частиц углерода и/или частиц диоксида титана в структуру материала.

Ключевые слова: технический углерод, диоксид титана, экран электромагнитного излучения.

Введение

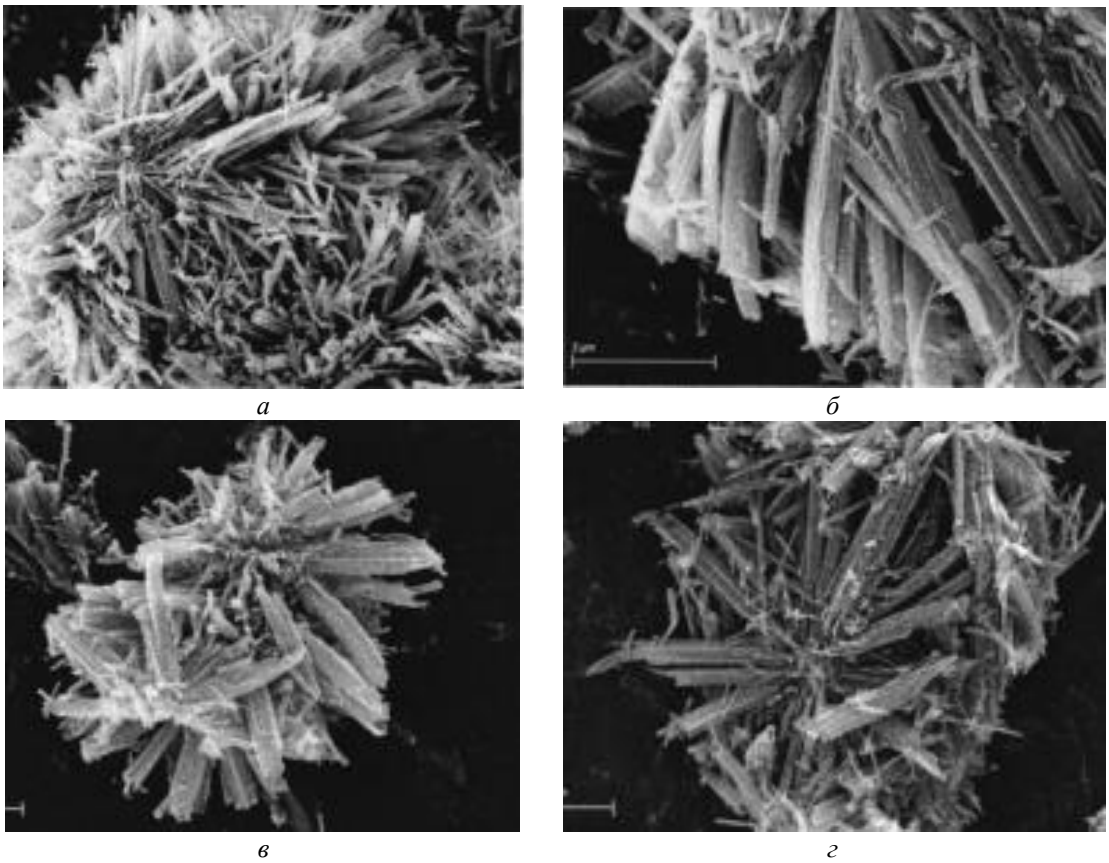
Данная статья является продолжением исследований, направленных на получение гибких экранов электромагнитного излучения, обладающих незначительными массогабаритными параметрами и эффективно ослабляющих электромагнитное излучение СВЧ-диапазона. В работе [1] представлены результаты исследования экранов электромагнитного излучения на основе волокнистого материала, содержащего частицы технического углерода, который можно использовать для создания изделий, способных защищать организм человека от негативного влияния электромагнитного излучения. В данной работе поставлена задача создания композита из диоксида титана и технического углерода и его внедрения в структуру синтетического материала.

Известно, что карбид титана производится из смеси порошка титана с техническим углеродом [2]. В промышленном производстве смесь, состоящая из 68,5 % TiO_2 и 31,5 % технического углерода, подвергается продолжительному и тщательному перемешиванию в жидкой среде, после чего прессуется под давлением в брикеты, которые загружаются в графитовые тигли. Карбидизация смеси проводится в атмосфере водорода в угольно-трубчатых печах или вакуумных садочных печах. В угольно-трубчатых печах графитовые тигли непрерывно продвигаются в печи, температура которой 2000 °С. Измельченный и просеянный карбид титана содержит 20–20,5 % углерода, из которых 2 % находится в виде свободного углерода. Определяющее влияние на состав получаемого по этому методу карбида титана оказывает равномерность смещения исходных компонентов. Равномерное смещение компонентов и высокая дисперсность диоксида титана и углерода достигаются с помощью растворного метода, который заключается в осаждении из солянокислого раствора трех-хлористого титана раствором аммиака гидрооксида титана. Гидрооксид титана после смешивания с водным раствором углеводов упаривают в потоке инертного газа.

Углерод в технологии высокотемпературного синтеза карбида титана берется обычно в виде сажи. В случае сухого смешивания исходных порошков титана и сажи большое влияние на качество смешивания оказывает влажность порошков. Увлажненные порошки слеживаются, теряют свойство сыпучести и плохо смешиваются. Кроме того, влага в порошках является источником кислорода, загрязняющего конечный продукт. Поэтому для удаления адсорбированной влаги порошки перед смешиванием необходимо просушивать.

Исследования, проведенные авторами в работе [3], позволили сделать вывод, что композит

диоксид титана/углерод (TiO_2/C) стержневой морфологии может быть получен простым и технологичным способом, без использования вредных и ядовитых компонентов при условии использования в процессе термического разложения прекурсора – глицеролата титана $Ti(C_3H_7O_3)_4$, что обеспечивает одновременное формирование в процессе термической деструкции прекурсора оксидной и углеродной составляющих композита. Кроме того, это позволяет получать композит TiO_2/C на основе разнообразных кристаллографических симметрий диоксида титана: аморфного диоксида титана, анатазной модификации, смешанных анатазной и рутильной модификаций. Согласно СЭМ, частицы композита TiO_2/C имеют морфологию стержней толщиной 20–250 нм и длиной до 4 мкм (рис. 1).



а *б* *в* *г*

Рис. 1. Изображение композита TiO_2/C стержневой морфологии, полученного в результате нагрева в течение 0,5 ч при температурах: *а* – 360 °C; *б* – 480 °C; *в* – 600 °C; *г* – 850 °C [3]

Методика создания образцов экранов электромагнитного излучения

Для создания образцов экранов электромагнитного излучения использовались порошки диоксида титана и технического углерода. Диоксид титана существует в виде нескольких кристаллических модификаций. В природе можно встретить анатаз, рутил и брукит. Следует отметить, что брукит промышленно почти не производится и в природе встречается редко. Анатазная форма также существенно уступает по производству рутильной, так как хуже рассеивает свет. Чистый диоксид титана представляет собой бесцветное кристаллическое вещество, желтеющее при нагревании. В тонкораздробленном состоянии – белый порошок. Практически не растворяется в воде и минеральных кислотах, кроме плавиковой и концентрированной серной. Температура плавления рутила: 1870 °C, температура кипения – 2500 °C. Плотность рутила при 20 °C составляет 4,235 г/см^3.

Технический углерод представляет собой высокодисперсный аморфный углеродный

продукт, производимый в промышленных масштабах. Порошкообразный технический углерод представляет собой набор частиц размером примерно 13–200 нм. Температура образования технического углерода составляет 1000 ºС и выше. С использованием электрофильтров сформированные частицы улавливаются с последующей технологической операцией удаления гранул и уплотнений. Технический углерод широко используется для создания проводящих компонентов, а также в качестве дисперсного наполнителя в полимерном связующем.

В данной работе в качестве основы для создания экранов электромагнитного излучения применялось нетканое иглопробивное полотно, которое содержало 70 % полиэфирных волокон, 20 % полипропиленовых волокон и 10 % угольных волокон марки УГЦВ-1-Р. Толщина полотна составляла 4 мм, поверхностная плотность – 250 г/м^2.

На основе ранее проведенных исследований [4] было решено использовать для создания образцов раствор дистиллированной воды с техническим углеродом и/или порошком диоксида титана. Методика пропитки волокнистых основ включала следующие этапы:

– подготовка порошков (помол, промывка, сушка в течение 24 ч);

– определение массы сухого порошкообразного материала и жидкостного наполнителя;

– приготовление водного углеродосодержащего раствора (помещение порошкообразного материала и жидкостного наполнителя в смешивающий механизм);

– раскрой волокнистого материала на фрагменты необходимого размера;

– помещение волокнистого материала в герметичную емкость с углеродосодержащим раствором на 2 ч;

– просушка пропитанного материала в сушильном шкафу при температуре 50 ºС в течение 1 ч.

По представленной методике были изготовлены следующие образцы:

– образец экрана электромагнитного излучения на основе синтетического материала с содержанием диоксида титана (образец № 1);

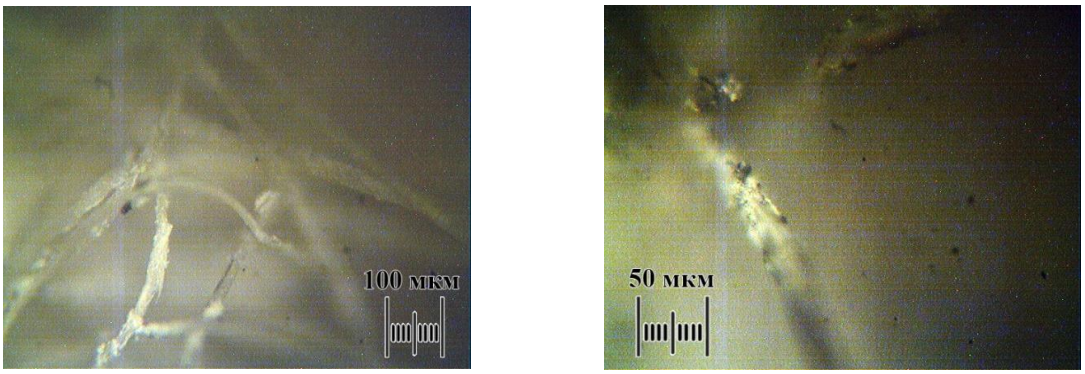
– образец экрана электромагнитного излучения на основе синтетического материала с содержанием технического углерода (образец № 2);

– образец экрана электромагнитного излучения на основе синтетического материала с содержанием композита TiO_2/C (образец № 3).

Для измерения коэффициентов передачи и отражения конструкций экранов электромагнитного излучения использовался панорамный измеритель коэффициентов передачи и отражения SNA 0,01–18, работающий по принципу раздельного выделения и непосредственного детектирования уровней падающей и отраженной волн. В состав панорамного измерителя входят генератор качающейся частоты (ГКЧ), блок обработки измерительных сигналов, передающая и приемная антенны, блоки направленных ответвителей, предназначенные для выделения и детектирования падающей, отраженной и прошедшей электромагнитных волн и соединяющиеся с каналами блока обработки измерительных сигналов и антеннами. Рабочий диапазон частот панорамного измерителя – 0,01...18 ГГц.

Для получения цифровых фотографий поверхности образцов экранов были использованы микроскоп МЕТАМ-Р1 и цифровая камера ЦК–13, которая предназначена для регистрации изображения и его записи на металлографических микроскопах.

На рис. 2 представлены микрофотографии поверхности образцов экранов электромагнитного излучения на основе синтетического материала с содержанием диоксида титана и композита TiO_2/C. Как видно из рис. 2, *а*, частицы диоксида титана закрепляются на синтетических волокнах и образуют дополнительный слой. То же самое происходит при пропитке синтетического полотна водным растровом с содержанием технического углерода. При использовании композита TiO_2/C на волокнах образуются слой, содержащий соединения частиц данных веществ (рис. 2, *б*).

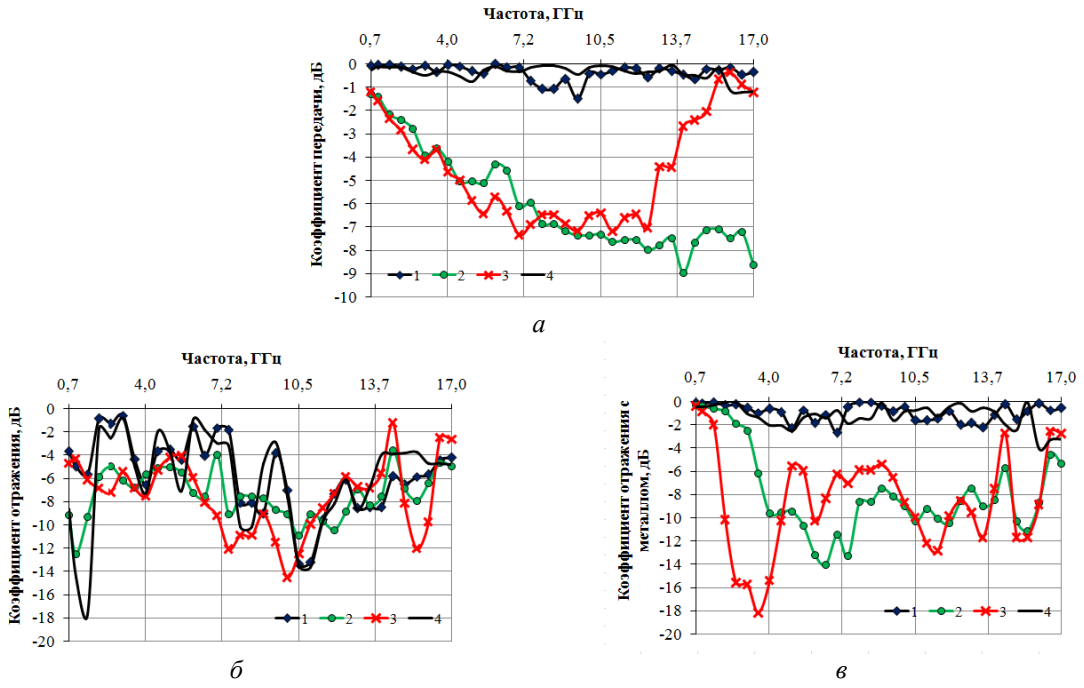


а *б*

Рис. 2. Микрофотографии поверхностей образцов экранов электромагнитного излучения на основе синтетического материала с содержанием: диоксида титана (*а*); и композита TiO_2/C (*б*)

Результаты проведенных исследований

Измерения коэффициентов отражения производились в режимах согласованной нагрузки и короткого замыкания (с установленным металлическим отражателем за образцом) в диапазоне частот 0,7–17 ГГц. Проведение измерений в данном диапазоне частот обусловлено тем, что в нем организовано функционирование систем мобильной связи, радиолокационных станций. Кроме того, в этом диапазоне лежат астоты информативных побочных электромагнитных излучений средств вычислительной техники. На рис. 3 представлены частотные зависимости коэффициентов передачи и отражения, измеренных в режиме согласованной нагрузки и короткого замыкания, для образцов экранов электромагнитного излучения на основе синтетического материала с содержанием диоксида титана (кривая 1), с содержанием технического углерода (кривая 2), с содержанием композита TiO_2/C (кривая 3), а также исходного синтетического материала (кривая 4).



а

б *в*

Рис. 3. Частотные зависимости коэффициентов: *а* – передачи; *б* – отражения, измеренного в режиме согласованной нагрузки; *в* – отражения, измеренного в режиме короткого замыкания

Из рис 3 следует, что введение порошка технического углерода в состав синтетического материала снижает коэффициенты отражения и передачи последнего. Коэффициент передачи образца № 2 (с содержанием технического углерода) составляет –1… –9 дБ, коэффициент

отражения изменяется в пределах от –3 дБ до –12 дБ (при измерении в режиме согласованной нагрузки) и от –1 дБ до –14 дБ (при измерении в режиме короткого замыкания). Частотные характеристики образца № 1 (с содержанием диоксида титана) незначительно отличаются от частотных характеристик исходного синтетического материала. При этом введение в состав синтетического материала композита из TiO_2/C в диапазоне частот 2–5 ГГц приводит к уменьшению коэффициента отражения до значения –18 дБ, измеренного в режиме короткого замыкания. При этом коэффициент передачи в данном диапазоне частот практически не изменяется и составляет –4 дБ.

Заключение

Таким образом, можно сделать вывод о том, что введение в состав волокнистого синтетического материала частиц технического углерода и диоксида титана увеличивает эффективность ослабления мощности электромагнитных волн в диапазоне частот 2–5 ГГц за счет образования слоя композита диоксида титана и технического углерода вокруг каждого волокна. Представленная структура представляет собой хаотическое расположение синтетических волокон, покрытых слоем композита из соединенных частиц диоксида титана и технического углерода. При этом необходимо отметить, что частицы не закреплены на волокнах, что может привести к их перемещению в следствие механической деформации, а значит и к изменению частотных характеристик коэффициентов отражения и передачи. В связи с этим дальнейшие исследования будут направлены на поиск способов закрепления частиц в структуре материала.

TITANIUM DIOXIDE/CARBON COMPOSIT FOR CREATING ELECTROMAGNETIC RADIATION SCREENS

M.S.Kh. Al-MAHDAWI, E.S. BELOUSOVA

Abstract. Studies of the intrusion processes of carbon and titanium dioxide particles into the synthetic material are presented. It is shown the changes in the frequency characteristics of the transmission and reflection coefficients measured in the modes of matched load and short circuit with the intrusion of carbon and/or titanium dioxide particles into the structure of the material.

Keywords: carbon black, titanium dioxide, electromagnetic radiation shield.

Список литературы

1. Аль-Махдави М.С.Х., Белоусова Е.С. // Докл. БГУИР. 2018. № 7 (117). С. 7–11.
2. Кипарисов С.С., Левинский Ю.В., Петров А.П. Карбид титана. Получение, свойства, применение. Москва, 1987.
3. Способ получения композита диоксид титана/углерод: пат. 2602536 Рос. Федерация. /Опубл. 20.11.2016.
4. Белоусова Е.С., Мохамед А.М.А., Аль-Адеми Я.Т.А. // Докл. БГУИР. 2017. № 2 (104). С. 63–68.

УДК 621.383

ОЦЕНКА НАДЕЖНОСТИ СЧЕТЧИКА ФОТОНОВ С МЕРТВЫМ ВРЕМЕНЕМ НА ОСНОВЕ АНАЛИЗА ВЕРОЯТНОСТИ ОШИБОЧНОЙ РЕГИСТРАЦИИ СИМВОЛОВ «0» В КВАНТОВО-КРИПТОГРАФИЧЕСКОМ КАНАЛЕ СВЯЗИ

А.М. ТИМОФЕЕВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 марта 2019

Аннотация. Выполнена оценка надежности счетчика фотонов с мертвым временем продлевающего типа с использованием в качестве критерия вероятности ошибочной регистрации символов «0». Применительно к квантово-криптографическому каналу связи установлено, что с увеличением средней скорости счета сигнальных импульсов на выходе счетчика фотонов вероятность ошибочной регистрации символов «0» вначале уменьшается, после чего растет. Причем рост средней длительности мертвого времени продлевающегося типа приводит к увеличению средних скоростей счета сигнальных импульсов на выходе счетчика фотонов, при которых достигаются наименьшие значения вероятности ошибочной регистрации символов «0».

Ключевые слова: счетчик фотонов, мертвое время, канал связи.

Введение

При обеспечении защиты инфокоммуникационных систем и сетей в настоящее время весьма часто используются квантово-криптографические каналы связи [1, 2]. Такие каналы связи характеризуются высоким уровнем информационной безопасности, т. к. позволяют распределять секретные криптографические ключи, шифровать и расшифровывать пользовательские данные, выполнять взаимную идентификацию и аутентификацию как пользователей, так и данных и др. [1–5]. Как известно, при создании каналов связи необходимо обеспечивать достаточно высокую надежность оборудования легитимных пользователей, что особенно актуально для квантово-криптографических каналов связи [6, 7].

Под надежностью будем понимать свойство оборудования выполнять возложенные на него функции информационной безопасности с сохранением своих характеристик (параметров) в определенных пределах при данных условиях эксплуатации.

Одним из критериев надежности является вероятность ошибочной регистрации данных [6]. Отметим, что ошибки в квантово-криптографических каналах связи во многом обусловлены тем, что передача информации осуществляется посредством маломощных оптических сигналов, содержащих в среднем от одного до нескольких десятков фотонов на каждый передаваемый бит (символ) [1–5]. Для регистрации таких сигналов в настоящее время преимущественно используются высокочувствительные приемные модули – счетчики фотонов, которые, однако, не способны регистрировать падающее на них оптическое излучение до истечения длительности их мертвого времени после регистрации бита (символа) [4, 8]. Поскольку до настоящего времени оценка влияния мертвого времени счетчика фотонов на вероятность ошибочной регистрации данных не выполнялась, это являлось целью данной работы.

Объектом исследования являлся асинхронный дискретный двоичный квантово-криптографический канал связи, в котором в качестве приемного модуля использовался счетчик фотонов с мертвым временем продлевающегося типа. Выбор в качестве объекта исследования такого канала связи объясняется тем, что в ряде случаев его использование оказывается более предпочтительным, ввиду отсутствия дополнительных линий связи для передачи и приема

синхроимпульсов [8]. Мертвым временем продлевающегося типа характеризуются счетчики фотонов на базе лавинных фотоприемников, включенных по схеме пассивного гашения лавины [1, 8].

Предметом исследования являлось установление влияния продлевающегося мертвого времени счетчика фотонов на вероятность ошибочной регистрации символов «0».

Выражение для оценки вероятности ошибочной регистрации символов «0»

Вначале получим выражение для расчета вероятности ошибочной регистрации символов «0», передаваемых по квантово-криптографическому каналу связи. Дальнейшие рассуждения будут основаны на том, что передача информации осуществляется с использованием дискретного двоичного асинхронного однородного квантово-криптографического канала связи без памяти и со стиранием [4, 9, 10]. Всеми потерями информации, за исключением потерь в счетчике фотонов, пренебрегаем.

На основании выражений для оценки вероятности ошибочной регистрации данных и статистических распределений, полученных в работе [9], применительно к счетчикам фотонов с рассматриваемым типом мертвого времени запишем выражение для вероятности ошибочной регистрации символов «0»:

$$P_{ош0} = 1 - \sum_{N=N_1}^{N_2} \frac{\left[\left(n_t + n_{s0}\right)\left(\Delta t - \tau_d\right)\right]^N \exp\left[-\left(n_t + n_{s0}\right)\left(\Delta t - \tau_d\right)\right]}{N!},$$

где N_1 и N_2 – нижний и верхний пороговые уровни регистрации соответственно, n_t – средняя скорость счета темновых импульсов на выходе счетчика фотонов, n_{s0} – средняя скорость счета сигнальных импульсов на выходе счетчика фотонов при передаче символов «0», Δt – среднее время однофотонной передачи, τ_d – средняя длительность мертвого времени продлевающегося типа.

Отметим, что для оценки мертвого времени продлевающегося типа используют среднее значение, т. к. его длительность зависит от интенсивности оптического излучения [8].

Темновые и сигнальные – это импульсы, которые появляются на выходе счетчика фотонов соответственно в отсутствии оптического сигнала и в результате воздействия фотонов регистрируемого излучения [8].

Нижний и верхний пороговые уровни регистрации – это соответственно наименьшее и наибольшее число зарегистрированных на выходе счетчика фотонов импульсов, при котором делается вывод, что передан символ «0». При превышении зарегистрированных импульсов числа N_2 делается вывод, что передан символ «1», а при регистрации импульсов в количестве, меньшем, чем N_1, принимается решение, что символ отсутствует [4, 9].

Приведенное выше выражение для оценки вероятности ошибочной регистрации символов «0» получено из следующих соображений. При подаче на вход счетчика фотонов регистрируемого излучения на его выходе формируется смесь темновых и сигнальных импульсов. Статистические распределения этих импульсов при наличии на входе счетчика фотонов ослабленного оптического излучения соответствуют распределению Пуассона [1, 8] и определяют выбор нижнего и верхнего пороговых уровней регистрации N_1 и N_2. Причем для рассматриваемого канала связи при передаче символов «0» и «1» используются оптические сигналы мощностью P_1 и P_2 ($P_1 < P_2$), которые транслируются в течение длительности Δt [4, 9]. Поскольку символы «0» и «1» передаются импульсами различной мощности, то на выходе счетчика фотонов за время Δt формируется различное количество электрических импульсов, которое будет прямо пропорционально мощности оптического излучения. Поэтому число импульсов, соответствующее символу «0», будет меньше, чем число импульсов, соответствующее символу «1» [4, 8, 9]. При этом вероятность $P_{ош0}$ имеет две составляющие. Первая составляющая определяет вероятность того, что при приеме оптического излучения счетчиком фотонов будет зарегистрировано меньше импульсов, чем установлено нижним пороговым уровенем, а вторая – вероятность того, что при наличии на входе счетчика фотонов оптического сигнала мощностью P_1 на его выходе будет зарегистрировано импульсов больше, чем установлено верхним пороговым уровенем регистрации [9–11]. Вероятность регистрации символов «0» при наличии на входе канала связи символов «0» равна $1 - P_{ош0}$.

На основании представленных рассуждений можно сделать вывод, что приведенное выражение пригодно для определения вероятности ошибочной регистрации символов «0» для рассматриваемого квантово-криптографического канала связи при соблюдении указанных выше ограничений.

Результаты и их обсуждение

Вычисление вероятности ошибочной регистрации двоичных символов «0» выполнялось для каналов связи, содержащих в качестве приемного модуля счетчик фотонов с мертвым временем продлевающегося типа при различных значениях τ_d и n_{s0}.

На рисунке представлены зависимости вероятностей ошибочной регистрации двоичных символов «0» от средней скорости счета сигнальных импульсов n_{s0} для различной средней длительности мертвого времени продлевающегося типа.

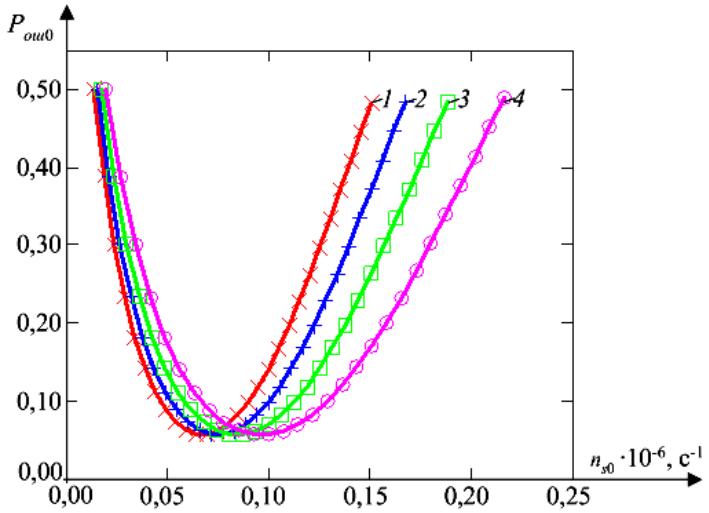


Рис. Зависимость вероятности ошибочной регистрации символа «0» от средней скорости счета сигнальных импульсов n_{s0} при средней длительности мертвого времени:
1 – × τ_d = 0; 2 – + τ_d = 5 мкс; 3 – □ τ_d = 10 мкс; 4 – ○ τ_d = 15 мкс

Зависимости $P_{ош0}(n_{s0})$ построены в диапазонах средних скоростей счета сигнальных импульсов, на которых вероятности регистрации на выходе канала связи символов «0» при наличии на входе канала связи символов «0» $P(0/0)$ составляют не менее 0,5 при заданных средних длительностях мертвого времени продлевающегося типа. Это обусловлено тем, что для рассматриваемого канала связи при $P(0/0) < 0{,}5$ использование счетчиков фотонов для регистрации данных становится нецелесообразным. Оценка переходных вероятностей $P(0/0)$ для рассматриваемого канала связи выполнялась по методике [11].

Расчет проводился для одинаковых значений нижнего и верхнего пороговых уровней регистрации $N_1 = 1$ и $N_2 = 7$, средней скорости счета темновых импульсов $n_t = 10^3$ с$^{-1}$ и среднего времени передачи одного бита (символа) τ_b = 100 мкс. Необходимо отметить, что пороговые уровни регистрации можно выбирать и другими, отличными от 1 и 7, но при сравнении значений $P_{ош0}(n_{s0})$ для различных средних длительностей мертвого времени следует фиксировать N_1 и N_2 постоянными, как и среднее значение скорости счета темновых импульсов n_t и среднее время передачи одного бита (символа) τ_b. При этом важно учитывать, что τ_d не может превышать Δt, которое, в свою очередь, должно быть меньше средней длительности передачи одного бита (символа) τ_b на величину защитного временного интервала (см. работы [4, 9]); в противном случае использование счетчиков фотонов для регистрации данных становится невозможным. Отметим, что при других значениях N_1 и N_2, и отношениях $\tau_d/\Delta t$ и n_t/n_{s0} проявление эффекта мертвого времени продлевающегося типа для рассматриваемого канала связи аналогично представленному на рисунке.

Из полученных графиков следует, что с увеличением n_{s0} зависимости $P_{ош0}(n_{s0})$ вначале спадают, достигая наименьшего значения $P_{ош0} = 0{,}06$, однако при дальнейшем увеличении n_{s0} –

растут. Причем рост средней длительности мертвого времени продлевающегося типа приводит к увеличению средних скоростей счета сигнальных импульсов n_{s0}, при которых достигаются наименьшие значения $P_{ош0}$. Так, например, наименьшие значения $P_{ош0}$ достигаются при $n_{s0} = 66{,}6 \times 10^3$ с$^{-1}$ для $\tau_d = 0$; при $n_{s0} = 74{,}1 \times 10^3$ с$^{-1}$ для $\tau_d = 5$ мкс; при $n_{s0} = 83{,}5 \times 10^3$ с$^{-1}$ для $\tau_d = 10$ мкс; при $n_{s0} = 95{,}6 \times 10^3$ с$^{-1}$ для $\tau_d = 15$ мкс.

Такое поведение зависимостей $P_{ош0}(n_{s0})$ объясняется смещением статистических распределений смеси числа темновых и сигнальных импульсов $P_{st0}(N)$ при передаче символов «0» с изменением средних скоростей счета сигнальных импульсов и мертвого времени продлевающегося типа, что достаточно подробно исследовано автором ранее (см. [10]). Распределения $P_{st0}(N)$ имеют явно выраженный максимум, свойственный распределению Пуассона, который с увеличением средней скорости счета сигнальных импульсов n_{s0} сдвигается в сторону больших значений N, как при наличии мертвого времени продлевающегося типа, так при его отсутствии. При $n_{s0} = 0$ максимум распределения $P_{st0}(N)$ соответствует значению $N = 0$, поэтому вероятность отсутствия импульсов на выходе счетчика фотонов достаточно большая, что определяет высокую вероятность ошибочной регистрации символов «0» (см. рисунок). С увеличением n_{s0} вероятность регистрации импульсов в количестве $N_1 \div N_2$ растет за счет сдвига $P_{st0}(N)$ в сторону больших значений N, поэтому вероятность ошибочной регистрации символов «0» уменьшается, и зависимости $P_{ош0}(n_{s0})$ спадают, достигая наименьшего значения (см. рисунок). При дальнейшем увеличении средней скорости счета сигнальных импульсов n_{s0} максимумы статистических распределений $P_{st0}(N)$ продолжают сдвигаться в сторону еще больших значений N. В результате увеличивается вероятность того, что на выходе счетчика фотонов будет зарегистрировано импульсов в количестве, превышающем верхний пороговый уровень регистрации N_2, поэтому $P_{ош0}$ растет (см. рисунок).

Также из рисунка видно, что в диапазонах средних скоростей счета сигнальных импульсов n_{s0}, на которых зависимости $P_{ош0}(n_{s0})$ уменьшаются, увеличение средней длительности мертвого времени продлевающегося типа при прочих равных параметрах приводит к росту вероятностей ошибочной регистрации символов «0». Однако в диапазонах n_{s0}, на которых зависимости $P_{ош0}(n_{s0})$ растут, увеличение τ_d, напротив, приводит к уменьшению $P_{ош0}$. Так, например, при $n_{s0} = 58{,}5 \times 10^3$ с$^{-1}$ и $n_{s0} = 107{,}8 \times 10^3$ с$^{-1}$ вероятности ошибочной регистрации символов «0» равны соответственно $6{,}24 \times 10^{-2}$ и $18{,}75 \times 10^{-2}$ для $\tau_d = 0$; $7{,}51 \times 10^{-2}$ и $13{,}02 \times 10^{-2}$ для $\tau_d = 5$ мкс; $9{,}57 \times 10^{-2}$ и $8{,}78 \times 10^{-2}$ для $\tau_d = 10$ мкс; $12{,}60 \times 10^{-2}$ и $6{,}27 \times 10^{-2}$ для $\tau_d = 15$ мкс. Объясняется это тем, что при увеличении τ_d максимумы статистических распределений $P_{st0}(N)$ сдвигаются в сторону меньших значений N [10]. При малых значениях n_{s0} максимумы распределений $P_{st0}(N)$ наблюдаются при количестве зарегистрированных импульсов, близких к нижнему пороговому уровню регистрации N_1, поэтому при увеличении средней длительности мертвого времени продлевающегося типа вероятности $P_{ош0}$ растут. Однако при достаточно больших значениях n_{s0} максимумы распределений $P_{st0}(N)$ соответствуют количеству зарегистрированных импульсов, превышающему верхний пороговый уровень регистрации N_2, поэтому с ростом τ_d вероятности ошибочной регистрации символов «0» уменьшаются.

Заключение

Получены зависимости вероятности ошибочной регистрации двоичных символов «0» $P_{ош0}$ от средней скорости счета сигнальных импульсов на выходе счетчика фотонов n_{s0} для различной средней длительности мертвого времени продлевающегося типа τ_d.

Определено, что с увеличением средней скорости счета сигнальных импульсов на выходе счетчика фотонов зависимости $P_{ош0}(n_{s0})$ спадают, достигая наименьшего значения, однако при дальнейшем увеличении n_{s0} – растут.

Установлено, что в диапазонах средних скоростей счета сигнальных импульсов n_{s0}, на которых зависимости $P_{ош0}(n_{s0})$ уменьшаются, увеличение средней длительности мертвого времени продлевающегося типа при прочих равных параметрах приводит к росту вероятностей ошибочной регистрации символов «0». Однако в диапазонах n_{s0}, на которых зависимости $P_{ош0}(n_{s0})$ растут, увеличение τ_d, напротив, приводит к уменьшению $P_{ош0}$.

Результаты, полученные в настоящей работе, могут быть использованы при создании систем квантово-криптографической связи, позволяющих с высокой достоверностью выявлять

несанкционированный доступ к каналу связи за счет уменьшения погрешности определения количества ошибок легитимного приемного оборудования, в качестве которого используются счетчики фотонов с мертвым временем продлевающегося типа.

ESTIMATION OF THE RELIABILITY OF THE PHOTON COUNTER WITH DEAD TIME BASED ON THE ANALYSIS OF THE PROBABILITY OF ERRONEOUS REGISTRATION OF SYMBOLS «0» IN A QUANTUM CRYPTOGRAPHIC COMMUNICATION CHANNEL

A.M. TIMOFEEV

Abstract. Reliability qualification photon counter with dead time prolonging type was performed. The criterion was the probability of erroneous registration of symbols «0». As applied to the quantum cryptographic communication channel, it has been established that with an increase in the average count rate of signal pulses at the output of the photon counter, the probability of erroneous recording of the symbols «0» first decreases and then increases. Moreover, the increase in the average duration of the dead time of the prolonged type leads to an increase in the average count rate of the signal pulses at the output of the photon counter, at which the smallest values of the probability of erroneous registration of the characters «0» are achieved.

Keywords: photon counter, dead time, communication channels.

Список литературы

1. Килин С.Я. Квантовая криптография: идеи и практика / С.Я. Килин; под ред. С.Я. Килин, Д.Б. Хорошко, А.П. Низовцев. – Минск: Белорус. наука, 2007.
2. Молотков С.Н. // Письма в ЖЭТФ. 2011. Т. 93. Вып. 3. С. 194–201.
3. Румянцев К.Е., Голубчиков Д.М. Квантовая связь и криптография: учебное пособие Таганрог: Изд-во ТТИ ЮФУ, 2009.
4. Тимофеев А.М. // Приборы и методы измерений. 2018. Т. 9. № 1. С. 17–27.
5. Румянцев К.Е., Пленкин А.П. // Известия ЮФУ. Технические науки. 2015. № 8 (169). С. 6–18.
6. Щеглов А.Ю. Анализ и проектирование защиты информационных систем. Контроль доступа к компьютерным ресурсам: методы, модели, технические решения СПб.: Профессиональная литература, 2017.
7. Молотков С.Н. // Письма в ЖЭТФ. 2005. Т. 81. Вып. 11. С. 733–738.
8. Гулаков И.Р., Зеневич А.О. Фотоприемники квантовых систем Минск: УО ВГКС, 2012.
9. Тимофеев А.М. // Актуальные проблемы науки XXI века. 2018. Вып. 7. С. 5–10.
10. Тимофеев А.М. // Вестник связи. 2018. № 1 (147). С. 56–62.
11. Тимофеев А.М. // Приборы и методы измерений. 2019. Т. 10. № 1. С. 80–89.

УДК 004.056(575.1)

НОРМАТИВНО-ПРАВОВЫЕ ВОПРОСЫ УКРЕПЛЕНИЯ КИБЕРБЕЗОПАСНОСТИ В РЕСПУБЛИКЕ УЗБЕКИСТАН

Х.К. САМАРОВ

Ташкентский университет информационных технологий имени Мухаммада-аль Хоразмий, Республика Узбекистан

Поступила в редакцию 20 марта 2019

Аннотация. В статье рассматриваются нормативно-правовые аспекты укрепления кибербезопасности в республике Узбекистан.

Ключевые слова: информационная система, информационный ресурс, информационная безопасность, кибербезопасность.

В свете развития информационных технологий все большую актуальность обретает проблема укрепления кибербезопасности. Данный вопрос неоднократно поднимается и главами государств. Так, Президент Республики Узбекистан в начале 2018 г. подписал Указ «О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций», в котором особое внимание уделяется реализации комплексных мер по обеспечению кибербезопасности и внедрению современных технологий защиты сетей, программных продуктов, информационных систем и ресурсов, участию в регулировании применения технологий сбора, обработки и хранения персональных и биометрических данных.

На сегодняшний день, ввиду стремительного распространения информационных сетей, актуальным является вопрос обеспечения надлежащего уровня кибербезопасности. При этом в условиях наличия активной угрозы нарушения прав и свобод граждан и организаций государству необходимо постоянно совершенствовать организационные, правовые и инженерно-технические механизмы контроля в отношении сетей телекоммуникаций. Ситуация осложняется тем, что на сегодняшний день активное участие в кибератаках, незаконном сборе информации о государственных служащих, корпорациях и простых гражданах принимают государственные структуры развитых стран. При этом применяются и различные методы манипуляция людьми при помощи сети Интернет [1]. Опыт последних лет указывает, что киберпреступность перешагнула границы отдельных государств и стала международной.

Лидерство в киберпространстве является и одним из приоритетов США. Американская национальная стратегия безопасности строится на дипломатии, разведке, военном комплексе, правоохранительных органах, а также экономических инструментах [2].

Однако самые эффективные методы защиты информационных данных и обеспечения кибербезопасности существуют в КНР. С 2003 г. в КНР действует проект «Золотой щит», который также называют «Великим китайским файерволом», так как большая часть его задач заключается в блокировке ненадежных сайтов и IP-адресов. Доступ к иностранным сайтам изнутри материкового Китая ограничивается правительством Китая. Веб-страницы фильтруются по ключевым словам, связанным с государственной безопасностью, а также по «черному списку» адресов сайтов. К инструментам работы «Золотого щита» можно отнести блокировку по IP адресу, DNS, URL, TCP фильтры, блокираторы подключения, создание фальшивых SSL подключений.

Закон КНР «О кибербезопасности» от 7 ноября 2016 г. (далее – Закон КНР 2016 года) предусматривает проведение ежегодных оценок рисков кибербезопасности и предоставление отчетов о результатах этих оценок и мерах по улучшению ситуации соответствующим органам власти (статья 38). Цель контроля над сетью Интернет в КНР – предотвратить проникновение нежелательной информации внутрь страны и утечку информации за рубеж, в том числе путем блокирования информационных ресурсов и поисковых систем [3].

Впрочем, опыт КНР в сфере обеспечения информационной безопасности часто подвергается критике, и является наиболее радикальным. Фактически специалисты КНР оградили свою часть глобальной сети «железным щитом» путем фильтрации любой информации, которая как поступает к ним, так и исходит от них.

Следует отметить, что, несмотря на законодательное урегулирование в Республике Узбекистан вопроса обеспечения кибербезопасности, эффективному процессу построения системы защиты информационного пространства препятствует ряд проблемных вопросов, требующих государственного урегулирования.

К таким вопросам можно отнести вопросы контроля над сетями телекоммуникаций, многие из которых находятся в частной собственности, вопросы защиты секретной информации, а также личных данных граждан Узбекистана, уязвимости в отечественных системах защиты информационных ресурсов.

Конвенция Совета Европы о киберпреступности разделила киберпреступность на четыре группы (дополнительный протокол добавил пятую группу) [4]. На сегодняшний день ряд государств успешно проводит политику укрепления информационной безопасности.

Исходя из международного опыта можно выделить три основные модели правового регулирования распространения информации в сети Интернет [5]. Первая модель предусматривает полный контроль государством над сетью Интернет. Данной модели придерживается, к примеру, КНР, где практически вся сеть Интернет находится под полным государственным контролем.

Вторая модель предусматривает ответственность провайдера за любые действия пользователя. Третья модель регулирования безопасности в сети Интернет предусматривает освобождение провайдера от ответственности в тех случаях, если он выполняет определенные условия, связанные с характером предоставления услуг и взаимодействия с субъектами информационного обмена.

Ключевым вопросом обеспечения информационной безопасности Республики Узбекистан является вопрос контроля над хранением и распространением информации. Информация, распространяемая в сетях телекоммуникаций, включает в себя как персональные данные пользователей, так и государственные секреты и закрытую информацию. Поэтому так важно иметь достаточную правовую основу для создания стабильной системы в сфере хранения и распространения информации.

Анализ законодательства в сфере хранения и распространения информации выявил ряд проблемных вопросов, которые требуют принятия мер для их последующего решения. Во-первых, на сегодняшний день источники угроз информационной безопасности могут находиться вне юрисдикции законодательства Республики Узбекистан, что существенно затрудняет применение системы правовых мер. Во-вторых, в соответствии с пунктом 1 статьи 18 Закона Республики Узбекистан «О телекоммуникациях», операторы связи и (или) владельцы сетей связи обязаны осуществлять содействие в сборе и хранении служебной информации в порядке, определяемом органами, проводящие оперативно розыскную деятельность. Однако закон не предписывает осуществлять сбор и хранение данных пользователей (переписка, звонки и т. д.). Для сравнения, в законе КНР «О мерах по регулированию информационных услуг через Интернет» от 1 октября 2000 г. предусмотрено, что провайдеры должны хранить такую информацию в течение времени, на которое его подписчики получают выход в Интернет, номера счетов подписчиков, адреса или названия доменов веб-сайтов и основные телефонные номера, которые они используют в течение 60 дней (статья 14). В-третьих согласно пункту 9 статьи 6 Закона Республики Узбекистан «Об информатизации» в компетенцию уполномоченного органа входит содействие по защите прав и законных интересов пользователей информационных ресурсов в вопросах безопасного использования информационно-коммуникационных технологий.

Для реализации данной компетенции уполномоченный орган обращается в суд с требованием блокировать те или иные информационные ресурсы, содержащие материалы, противоречащие законодательству Республики Узбекистан (пропаганда экстремизма, разжигание межнациональной розни, распространение порнографии и т. п.).

С 2012 г. в России действует Единый реестр запрещенных сайтов. Реестр – это конкретный набор из ста тысяч адресов. Реестр находится в ведении Роскомнадзора, в соответствии с постановлением Правительства Российской Федерации от 26 октября 2012 г. № 1101.

Необходимо рассмотреть возможность введения аналогичного механизма и в Республике Узбекистан. В целях повышения эффективности, упрощения процедуры и оперативного принятия мер по пресечению распространения незаконного контента предлагается следующее.

1. Упростить процедуры блокировки Интернет-ресурсов, путем возможности самостоятельной блокировки уполномоченным органом без обращения в суд. В качестве варианта можно перенять российский опыт ведения реестра запрещенных сайтов.

2. Распределить функции мониторинга Интернет-пространства между некоторыми государственными органами (например, анализ контента на предмет наличия угроз национальной безопасности отнести к введению органов службы государственной безопасности).

3. Сочетать при мониторинге как «ручные» способы поиска, так и автоматизированные.

Европейский Союз в рамках инициативы «Европа–2020» определил собственную Цифровую повестку дня (Digital Agenda) с обязательством выполнения широкого круга задач. Первая группа задач повестки ориентирована на дальнейшую популяризацию сети Интернет. Вторая группа задач сводится к обеспечению кибербезопасности граждан. На базе Европейского союза было создано Агентство по сетям и информационной безопасности (ENISA), которое постоянно проводит мониторинг мнений пользователей сети, в соответствии с этим вносит поправки в уже принятые проекты, которые становятся законом и соблюдаются странами ЕС. Отметим, что наиболее важное во всей этой инициативе – то, что законодатели должны проводить ежегодные встречи с политиками, IT-специалистами, учеными для обучения и совершенствования навыков безопасного использования сети Интернет, приучаясь тем самым к виртуальной культуре.

Другой проблемой в сфере безопасности информационных ресурсов является вопрос аттестации объектов информатизации. В соответствии постановлением Кабинета Министров Республики Узбекистан «О совершенствовании нормативно-правовой базы в сфере информатизации» от 22 ноября 2005 г. № 256, Приложение № 2, подпункт 28, примечание 1, обязательной аттестации подлежат: информационная система (ИС) государственного органа и Интернет-ресурс государственного органа. Автор предлагает усовершенствовать мероприятия по обеспечению кибербезопасности следующим образом.

1. Обязать пользователей сети Интернет осуществлять пользование социальными сетями, а также публикацию и комментирование записей и постов только под своими реальными именами.

2. Дополнить действующие Уголовный кодекс и Кодекс об административных правонарушениях Республики Узбекистан составами правонарушений, предусматривающими ответственность за неисполнение или ненадлежащее исполнение обязанностей по обеспечению безопасности государственных информационных систем.

3. Предусмотреть обязательный порядок аттестации объектов информатизации (информационные системы, программное обеспечение) не только для государственных, но и для частных информационных систем.

Подобный подход позволит осуществлять дополнительный контроль за соблюдением банками и субъектами государственного сектора при эксплуатации информационных систем единых требований в области информационно-коммуникационных технологий и обеспечения информационной безопасности.

Следует отметить, что одной из значительных мер, предпринятых Республикой Узбекистан в сфере противодействия киберпреступности, явилось включение в новый Уголовный кодекс и Кодекс об административных правонарушениях 2007 г. отдельных глав, посвященных правонарушениям в сфере информационных технологий. Таким образом, можно констатировать систематизацию правоприменительной деятельности в части противодействия правонарушениям в сфере информационных технологий.

LEGAL ISSUES OF CYBERSECURITY STRENGTHENING IN THE REPUBLIC OF UZBEKISTAN

Kh.K. SAMAROV

Abstract. The article deals with the regulatory and legal aspects of strengthening cybersecurity in the Republic of Uzbekistan.

Keywords: information system, information resource, information security, cybersecurity.

Список литературы

1. Безкоровайный М.М., Татузов А.Л. // Вопросы кибербезопасности. 2014. № 1 (2). С. 22–27.
2. Батуева Е.В. // Вестник МГИМО Университета. 2010. №4. С. 271–276.
3. Булавин А.В. // Общество: политика, экономика, право 2014. № 1. С. 27–31.
4. Номоконов В.А., Тропина Т.Л. // Криминология: вчера, сегодня, завтра. 2012. №24. С. 45–55.
5. Погорелова М.А. // Бизнес в законе. Экономико-юридический журнал. 2009. №2. С. 198–200.

УДК 621.392

МОДЕЛИРОВАНИЕ В СИМУЛЯТОРЕ NS-3 СЕТИ MANET НА ОСНОВЕ ПРОТОКОЛОВ МАРШРУТИЗАЦИИ AODV И DSDV

В.А. БЕЛАН, М.М. ШАКИР, М.Ю. ХОМЕНОК

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 18 марта 2019

Аннотация. Проанализированы системные характеристики самоорганизующейся сети с динамически изменяющейся топологией на базе симулятора NS-3. Выполнен сравнительный анализ протоколов с реактивной и проактивной маршрутизацией AODV и DSDV для двух моделей мобильности RWP и RWM.

Ключевые слова: сетевой симулятор, NS-3, NetAnim, моделирование.

Введение

Топология мобильной самоорганизующей сети определяется набором мобильных узлов, которые могут общаться без дополнительной инфраструктуры и централизованной организации маршрутов. Каждый узел в MANET работает как маршрутизатор, обеспечивая пересылку пакетов информации для других мобильных узлов.

Такие сети обладают рядом преимуществ:

– быстрая развертываемость из-за отсутствия необходимости строительства базовых станций;

– снижение требований к мощности передающих устройств мобильных узлов, т. к. дальность каждого должна быть больше или равна расстоянию до ближайшего узла, а не до базовой станции;

– возможность покрытия большой территории беспроводными маломощными узлами сети за счет их количества или обеспечение оптимального покрытия обслуживаемой территории путем эффективного занятия местоположения для покрытия максимально возможной площади;

– высокая отказоустойчивость (место потерянного узла может занять любой ближайший узел).

Однако особенность взаимодействия узлов в сети MANET делает трудным гарантировать качество услуг в условиях динамически изменяющейся ее топологии, определяемой различными моделями мобильности узлов сети. Поэтому к проблемным вопросам, требующим выполнения предварительных исследований перед развертыванием сети, является моделирование сетевой топологии с учетом мобильности узлов и оценка эффективности используемых протоколов маршрутизации имитационным моделированием сетевыми симуляторами.

Выделяют 3 класса протоколов маршрутизации: проактивные, реактивные и комбинированные (гибридные).

В проактивных протоколах при изменении топологии сети инициируется широковещательная рассылка сообщений об этих изменениях. При этом все маршруты хранятся в памяти каждого узла, и он может воспользоваться ими в любой момент. К ним относятся DSDV, TBRPF, FSR и OLSR.

В реактивных протоколах маршрутизации маршруты существуют только тогда, когда они необходимы. К протоколам с реактивной маршрутизацией относятся AODV, DSR, LMR и TORA.

Гибридные протоколы сочетают в себе подходы проактивных и реактивных протоколов на разных уровнях сетевой иерархии.

Существует множество моделей мобильности сети, с которыми может работать NS-3. К ним относятся:

– группа мобильности с контрольной точкой (Reference point group mobility);
– модель со случайной путевой точки (Random waypoint);
– Гаусовско-Марковская модель (Gauss-Markov);
– модель мобильности со случайным порядком (Random walk);
– модель Manhattan grid и др.

Основными параметрами моделей являются начальное местоположение мобильных узлов, их направление движения, диапазон скоростей, скорость изменяется со временем. Движение узлов в данных моделях описывается различными траекториями. Использование разных моделей мобильности и протоколов маршрутизации с одинаковыми параметрами симуляции может привезти к различным результатам качества обслуживания трафика.

Поэтому существует необходимость в изучении влияния моделей мобильности на производительность используемых протоколов в сети.

Визуализация сетевой топологии в аниматоре NetAnim

В качестве моделей мобильности были выбраны Random WayPoint Mobility Model (RWP) (рис. 1) и Random Walk Mobility Model (RWM) (рис. 2).

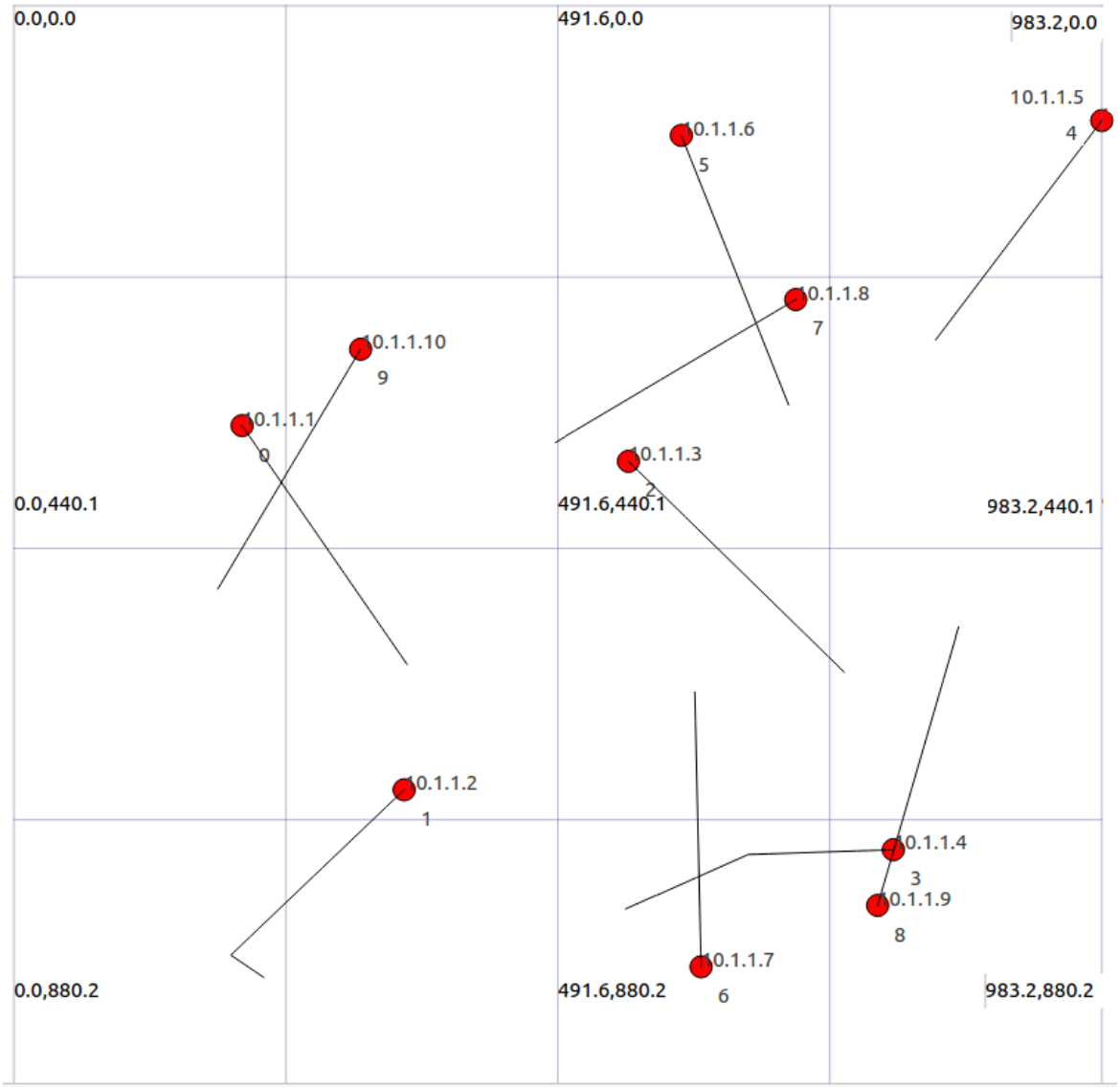


Рис. 1. Модель RWP

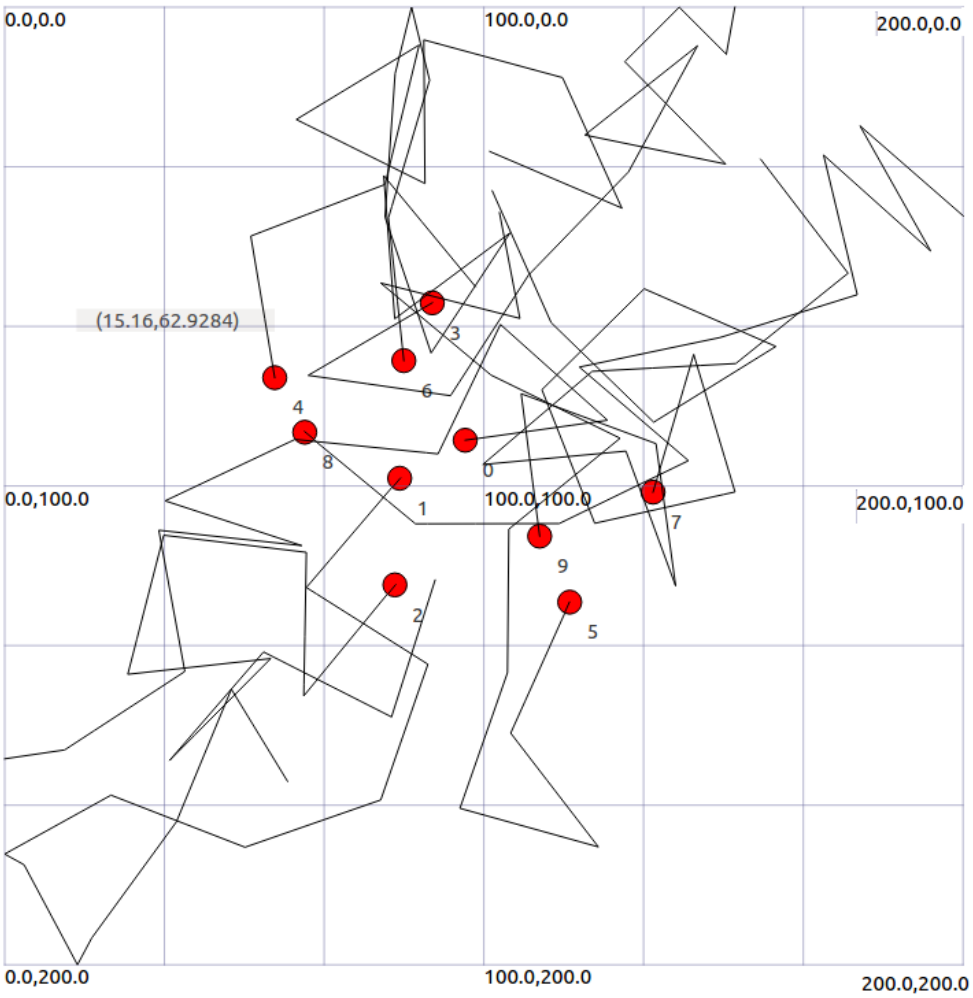


Рис. 2. Модель RWM

Для визуализации топологий сети в NS-3 используется модуль NetAnim (рис. 1). Модуль NetAnim, помимо визуализации топологии, позволяет выполнить пошаговую симуляцию, вывод таблиц маршрутизации во времени, просмотреть информацию о передаваемых пакетах и их трейсы взаимодействия, а также позволяет работать с модулем Flow Monitor, предоставляющим гибкие методы сбора самых различных показаний моделируемых активных сетевых устройств и каналов связи. Модуль позволяет полностью оценить характеристики сети и построить гистограммы для визуализации полученных данных. Оба модуля работают в XML файле [2].

В модели RWP предполагается фиксированное количество узлов в прямоугольнике фиксированного размера. Моделирование начинается с узлов, равномерно распределенных в прямоугольнике. Для каждого узла выбирается случайный пункт назначения и случайная скорость, равномерно распределенная в интервале $[v_{min}, v_{max}]$. Как только узел прибывает в пункт назначения, он останавливается на случайное время, равномерно распределенное в $[P_{min}, P_{max}]$, затем он выбирает новую скорость и пункт назначения, и процесс повторяется.

Модель RWM соответствует броуновскому движению, при котором каждый узел в случайном порядке выбирает направление, равномерно распределенное в интервале $[0, 2\pi]$, и скорость, равномерно распределенную в интервале $[v_{min}, v_{max}]$. Затем узел движется в течение определенного периода времени (или на фиксированное расстояние) с этой скоростью, после чего процесс повторяется.

Исходными параметрами для обеих моделей мобильности являются 10 узлов, 5 из которых – источники, а остальные – приемники, с установленными IP-адресами и с ID-нумерацией 0…9. На рис. 2 линиями показаны траектории движения узлов за время симуляции сети.

Узлы движутся со скоростью 10 м/с и характеризуются временем обновления 2 с. Узлы передают UDP-потоки с постоянной скоростью передач 8 кбит/с, что соответствует типу трафика CBR (Constant Bit Rate). Время симуляции проекта – 30 с.

На рис. 3 представлены таблицы маршрутизации узла № 1, динамика которого описывается моделью RWP, для AODV и DSDV протоколов маршрутизации соответственно.

Node: 1, Time: +29.0s, Local time: +29.0s, DSDV Routing table

DSDV Routing table

Destination	Gateway	Interface	HopCount	SeqNum	LifeTime	SettlingTime
10.1.1.1	10.1.1.1	10.1.1.2	1	12	3.905s	5.000s
10.1.1.3	10.1.1.1	10.1.1.2	3	12	3.934s	0.000s
10.1.1.4	10.1.1.1	10.1.1.2	2	12	3.934s	0.000s
10.1.1.5	10.1.1.1	10.1.1.2	3	12	3.905s	0.000s
10.1.1.6	10.1.1.1	10.1.1.2	3	10	8.933s	0.000s
10.1.1.7	10.1.1.1	10.1.1.2	2	12	3.934s	0.000s
10.1.1.8	10.1.1.10	10.1.1.2	2	12	3.961s	0.000s
10.1.1.9	10.1.1.1	10.1.1.2	4	12	3.931s	0.000s
10.1.1.10	10.1.1.10	10.1.1.2	1	12	3.924s	0.000s
10.1.1.255	10.1.1.255	10.1.1.2	0	12	-9223.855s	0.000s
127.0.0.1	127.0.0.1	127.0.0.1	0	0	-9223.855s	0.000s

Node: 1; Time: +29.0s, Local time: +29.0s, AODV Routing table

AODV Routing table

Destination	Gateway	Interface	Flag	Expire	Hop:
10.1.1.1	10.1.1.1	10.1.1.2	UP	2.95	1
10.1.1.3	10.1.1.1	10.1.1.2	UP	2.95	3
10.1.1.4	10.1.1.4	10.1.1.2	UP	2.99	1
10.1.1.5	10.1.1.1	10.1.1.2	UP	2.72	3
10.1.1.7	10.1.1.1	10.1.1.2	UP	2.41	2
10.1.1.8	10.1.1.1	10.1.1.2	UP	2.87	2
10.1.1.9	10.1.1.1	10.1.1.2	UP	2.48	3
10.1.1.10	10.1.1.10	10.1.1.2	UP	2.94	1
10.1.1.255	10.1.1.255	10.1.1.2	UP	92007.85	1
127.0.0.1	127.0.0.1	127.0.0.1	UP	92007.85	1

Рис. 3. Таблицы маршрутизации AODV и DSDV проколов
в момент времени окончания симуляции при RWP модели

Node: 1; Time: +29.0s, Local time: +29.0s, AODV Routing tab

AODV Routing table

Destination	Gateway	Interface	Flag	Expire	Hops
10.1.1.1	10.1.1.1	10.1.1.2	UP	2.57	1
10.1.1.3	10.1.1.3	10.1.1.2	UP	3.00	1
10.1.1.4	10.1.1.4	10.1.1.2	UP	2.66	1
10.1.1.5	10.1.1.5	10.1.1.2	UP	2.48	1
10.1.1.6	10.1.1.6	10.1.1.2	UP	2.34	1
10.1.1.7	10.1.1.7	10.1.1.2	UP	2.67	1
10.1.1.8	10.1.1.8	10.1.1.2	UP	2.22	1
10.1.1.9	10.1.1.9	10.1.1.2	UP	2.26	1
10.1.1.10	10.1.1.10	10.1.1.2	UP	2.81	1
10.1.1.255	10.1.1.255	10.1.1.2	UP	92007.85	1
127.0.0.1	127.0.0.1	127.0.0.1	UP	92007.85	1

Node: 1, Time: +29.0s, Local time: +29.0s, DSDV Routing table

DSDV Routing table

Destination	Gateway	Interface	HopCount	SeqNum	LifeTime	SettlingTime
10.1.1.1	10.1.1.9	10.1.1.2	2	12	3.911s	0.000s
10.1.1.3	10.1.1.3	10.1.1.2	1	12	3.908s	0.000s
10.1.1.4	10.1.1.4	10.1.1.2	1	12	3.899s	0.000s
10.1.1.5	10.1.1.5	10.1.1.2	1	12	3.899s	5.000s
10.1.1.6	10.1.1.6	10.1.1.2	1	12	3.897s	0.000s
10.1.1.7	10.1.1.7	10.1.1.2	1	12	3.896s	0.000s
10.1.1.8	10.1.1.8	10.1.1.2	1	12	3.900s	0.000s
10.1.1.9	10.1.1.9	10.1.1.2	1	12	3.898s	0.000s
10.1.1.10	10.1.1.10	10.1.1.2	1	12	3.898s	0.000s
10.1.1.255	10.1.1.255	10.1.1.2	0	12	-92007.855s	0.000s
127.0.0.1	127.0.0.1	127.0.0.1	0	0	-92007.855s	0.000s

Рис. 4. Таблицы маршрутизации AODV и DSDV проколов
в момент времени окончания симуляции при RWM модели

По таблицам маршрутизации можно определить адрес назначения, узел, через который направлен пакет и количество скачков до узла назначения, если нет прямой связи между узлами. Как следует из таблиц, реактивному протоколу AODV необходимо меньше скачков, чтобы достигнуть до узла назначения. Модели RWM соответствует меньшее количество скачков для обоих протоколов маршрутизации.

Результаты симуляции

Для вывода результатов симуляции проекта используется модуль Flow Monitor. Это модуль мониторинга потоков, который упрощает сбор и сохранение в постоянном хранилище общего набора показателей производительности сети. Модуль автоматически обнаруживает все потоки, проходящие через сеть, и сохраняет показатели, которые исследователю могут потребоваться для анализа потоков, например, битрейты, продолжительность передачи, задержки, размеры пакетов и коэффициент потери пакетов.

В процессе выполнения программы имитационного моделирования фрагмента сети Manet модулем Flow Monitor сформировано 45 потоков, а результаты статистической оценки системных параметров анализируемой сети представлены в виде гистограмм скоростей передачи

(Flow_bitrates), количества потерянных пакетов (Number_of_lost_packets) и задержки (Delay) по каждому направлению потоков (Number_of_flows).

На рис. 5 представлены результаты обработки данных для модели мобильности RWP.

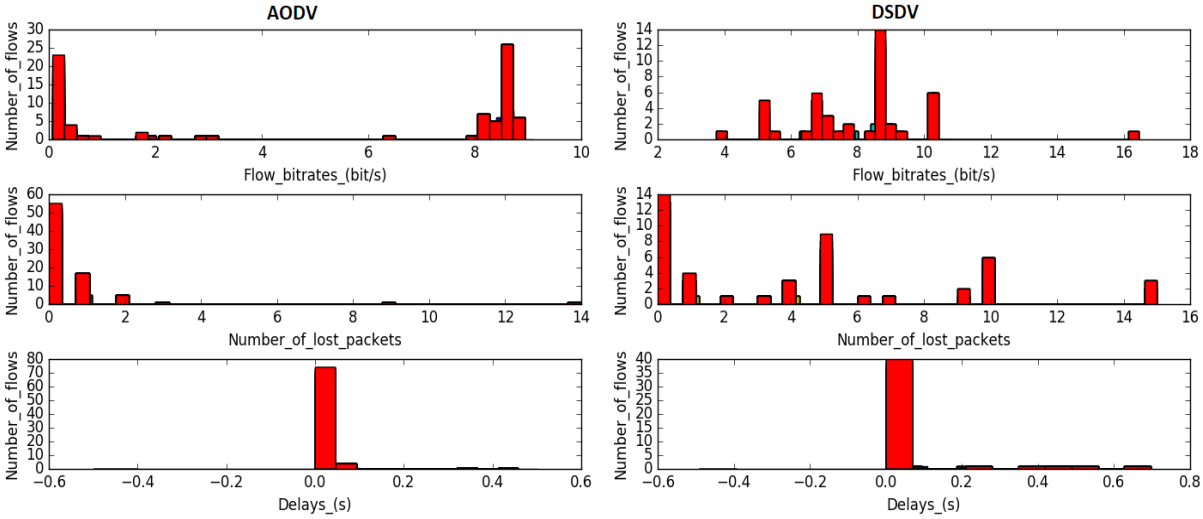


Рис. 5. Результаты статистической обработки для модели RWP по двум протоколам

Из гистограмм видно, что для проектируемой сети с моделью мобильности RWP реактивный протокол AODV показал наименьшие потери пакетов и задержки, протокол DSDV выделяется наибольшей скоростью передачи пакетов.

На рис. 6 представлены результаты обработки данных для модели мобильности RWM.

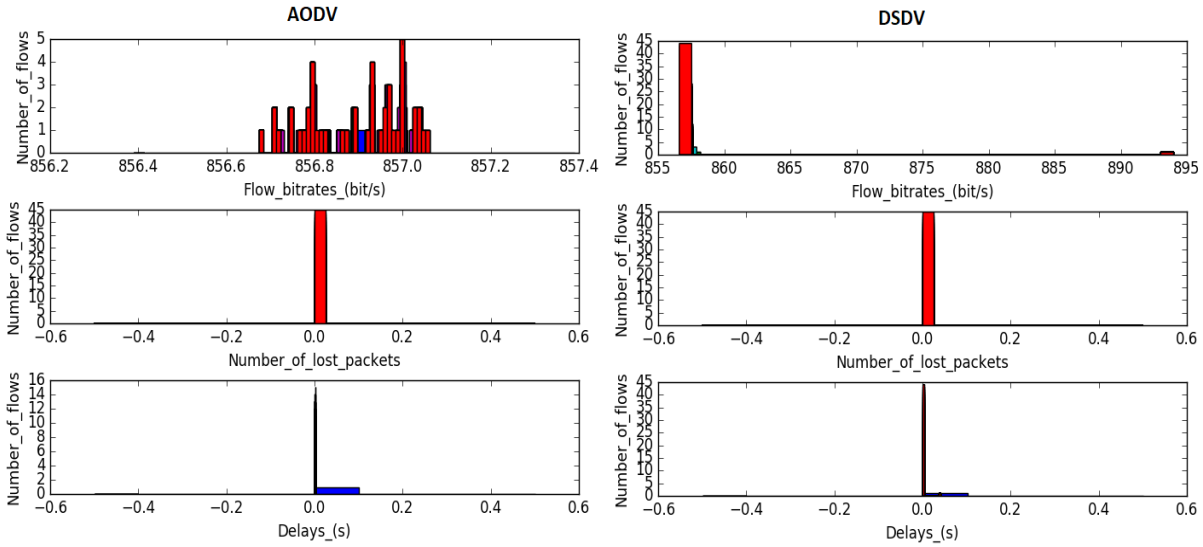


Рис. 6. Результаты статистической обработки для модели RWM по двум протоколам

Из гистограмм видно, что для проектируемой сети с моделью мобильности RWM реактивный протокол AODV и проактивный DSDV показали нулевые потери пакетов и минимальные задержки, по скорости передачи пакетов результаты по двум протоколам так же схожи, и в среднем составляют 857 бит/с.

В таблице сведены основные статистические данные имитационного моделирования сетевого фрагмента.

Результаты симуляции сети

	AODV		DSDV	
	RWP	RWM	RWP	RWM
Общее количество переданных пакетов	1516	1125	1125	1125
Общее количество полученных пакетов	1457	1125	838	1125
Общее количество потерянных пакетов	53	0	202	0
Пропускная способность, кб/с	5,61	8,36	6,77	8,36
Процент доставки пакетов, %	96	100	82	100
Средняя задержка, с	0,016	0,0012	0,064	0,0024
Средний джиттер, с	0,015	0,00027	0,05	0,0014
Размер пакета, кб	774	1028	1028	1028

Заключение

Для анализа качества функционирования сети на основе полученных статистических данных моделирования могут быть определены следующие основные метрики:

– пропускная способность (Throughput) как среднее количество пакетов, переданных успешно от одного узла к другому за время симуляции;

– коэффициент отброшенных пакетов (Packet Drop ratio), – как среднее количество пакетов, которые были потеряны между узлами, источником и назначения, к общему числу пакетов;

– коэффициент принятых пакетов (Packet Received ratio) как отношение среднего количества пакетов, которые были успешно переданы между узлами, источником и назначения, к общему числу пакетов.

Анализ результатов симуляции сетевого фрагмента показывает, что выходные данные существенно зависят от моделей мобильности узлов в сети. Модель с большей плотностью более устойчива к потере пакетов из-за разнообразия маршрутов до узла назначения.

Протоколы маршрутизация также влияют на результаты в зависимости от моделей мобильности. Так, реактивный протокол AODV реализует лучшие показатели при модели со случайной путевой точкой, но при этом характеризуется меньшей скоростью передачи по сравнению с проактивным протоколом DSDV. А при модели со случайным порядком оба протокола эквиваленты.

MODELING MANET NETWORK IN NS-3 SIMULATOR ON BASE OF AODV AND DSDV ROUTING PROTOCOLS

V.A. BELAN, M.M. SHAKIR, M.Yu. HOMENOK

Abstract. The system characteristics of a self-organizing network with a dynamically changing topology which use the Manet technology on base of the NS-3 simulator were analyzed. The comparative analysis both of proactive DSDV and reactive AODV routing protocols for two RWP and RWM mobility models was performed.

Keywords: modeling, NS-3, simulator, NetAnim.

Список литературы

1. NS-3 [Электронный ресурс]. URL: http://clone.nsnam.org/ns-3-allinone (дата обращения: 10.03.2019).
2. NetAnim [Электронный ресурс]. URL: https://www.nsnam.org/wiki/NetAnim (дата обращения: 10.03.2019).

УДК 621.391.63

ОЦЕНКА ВОЗМОЖНОСТИ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА К ПАССИВНОЙ ОПТИЧЕСКОЙ СЕТИ

Н.Н. СЕРГЕЕВ, В.Н. УРЯДОВ

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

Поступила в редакцию 20 марта 2019

Аннотация. Рассматривается возможность несанкционированного доступа к информации передаваемой от абонента, при использовании отраженного сигнала от ответвителя на другой абонентской розетке. Проведен расчет уровня полученного отраженного сигнала. а также расчет порога чувствительности приемника. Выполнена оценка возможности такой атаки.

Ключевые слова: пассивная оптическая сеть, защита информации, несанкционированный доступ, сетевая атака и угроза.

Введение

Недостатком PON-сети является незащищенность канала передачи от действий злоумышленников. В статье [1] показано, что для того чтобы воспользоваться отраженным сигналом и снять информацию абонента с другой соседней оптической розетки, достаточно, чтобы величина отраженного сигнала была в рамках квантового предела детектируемости. Однако чувствительность в квантовом пределе недостижима, и обычно применяют приемный оптический модуль с оптическим предусилителем.

Целью работы является исследование возможности несанкционированного доступа к информации, передаваемой от абонента, при использовании отраженного от ответвителя сигнала на другой абонентской розетке в пассивной оптической сети.

Вид рассматриваемой угрозы

На канальном уровне проблема безопасности в PON-сети практически решена, защита обеспечивается современными алгоритмами шифрования данных. На физическом – неисправность лазера обратного канала или контроллера этого лазера может вывести из строя всю систему. В принципе, такие случаи предусмотрены и отслеживаются системами управления, а включившийся на непрерывную передачу лазер может быть отключен с помощью watchdog, контролирующим, чтобы время функционирования лазера не превышало отведенный временной интервал. Куда серьезнее могут быть последствия действий злоумышленников: для нарушения (или даже прекращения) работы всего сегмента PON-сети достаточно осуществить воздействие лазером на любое из абонентских окончаний или просто подключить к нему любое активное сетевое устройство. Достаточно просто реализуемо и снятие информации нисходящего потока, поскольку обычный приемник обеспечивает прием сигнала любого приемника, если использовать другой временной интервал. Более сложно в реализации снятие информации контролируемого абонента с другой абонентской розетки, поскольку для этого необходимо использовать отраженный от ответвителя сигнал. При использовании отраженного сигнала необходимо учесть, что в пассивной оптической сети существуют возвратные потери в неоднородностях.

Расчет уровня полученного отраженного сигнала

Рассмотрим приведенную на рис. 1 модель, состоящую из станционного терминала OLT, оптической линии, разветвителя и нескольких абонентских терминалов ONT.

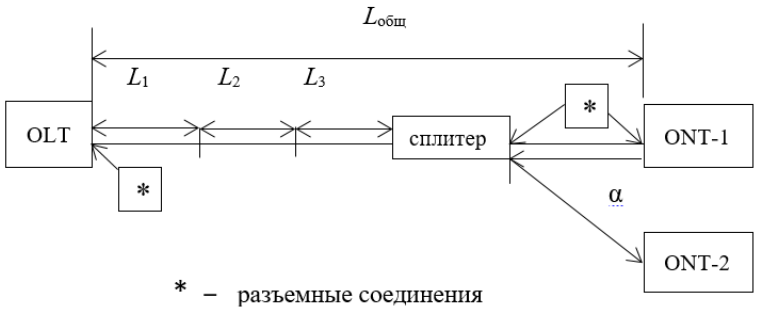


Рис. 1. Поток информации в PON-сети

Для снятия информации, передающейся от ONT-1 в восходящем (upstream) направлении, с ONT-2, будет использоваться отраженный от ответвителя сигнал $\alpha_{р.}$ Для каждой оптической линии представим все потери в линии в виде суммы затуханий всех компонентов: оптического кабеля, разъемных соединений, сварных соединений, разветвителей:

$$A = A_{отв} + A_{раз}, [дБ] \quad (1)$$

где A – суммарные потери, $A_{отв}$ – потери на оптическом разветвителе, $A_{раз}$ – потери в разъемном соединении. В рассматриваемом случае на пути прохождения сигнала ONT-1 – ONT-2 имеется 4 разъемных соединения, затухание в которых, согласно допустимым нормам [3], составит:

$$A_{раз} = A_{раз.норм} \cdot N_{раз} = 0,5 \cdot 4 = 2, [дБ] \quad (2)$$

Потери на оптическом ответвителе находим как возвратные потери, которые определяют долю оптической мощности, возвращающейся обратно к источнику оптического сигнала. В рассматриваемом случае источником оптического сигнала является ONT-1. Возвратные потери определяются согласно выражению:

$$\alpha_B = -10\lg\frac{P_{отр}}{P_{вх}} = -10\lg\frac{(n-n_c)^2}{n^2+n_c^{\,2}}, [дБ] \quad (3)$$

где $P_{вх}$ и $P_{отр}$ – мощности падающего и отраженного излучения, соответственно.

Основным фактором, определяющим возвратные потери, являются френелевские отражения. С этой позиции разветвитель характеризуется затуханием на ближнем конце 50 дБ [4]. В результате получается:

$$2 + 50 = 52, [дБ] \quad (4)$$

Получен уровень сигнала ONT-1, который придет на терминал ONT-2. Рассчитаем, достаточен ли будет полученный уровень сигнала, для снятия информации на ONT-2. Для этого найдем порог чувствительности приемника. Так как одним из основных модулей приема на сегодняшний день является оптический модуль с pin-фотодетектором, то расчет будет проведен для такого детектора.

Расчет порога чувствительности приемника

В работе [5] показано, что значение порога чувствительности для фотоприемника с pin-фотодетектором определяется следующим выражением:

$$P_{pin} = \frac{A_\lambda}{\eta_m}\sqrt{\overline{i^2}_{pin}} \quad (5)$$

где $A_\lambda = P_{ош}(hc/e\lambda)$ – коэффициент, пропорциональный энергии падающего фотона; $P_{ош}$ – параметр, характеризующий вероятность ошибки (в нашем случае $P_{ош}$=6,36, что соответствует $P_{ош} = 10^{-10}$); h, c, e – физические постоянные, h – постоянная Планка, c – скорость света, e – заряд электрона; η_m – квантовая эффективность – величина,

показывающая эффективность преобразования фотон-электрон для современных фотоприемников она определена и равна η_m =0,75–0,9; $\sqrt{\overline{i^2}_{pin}}$ – среднеквадратичное значение шумового тока приемного модуля с pin-фотодиодом.

При длине волны λ равной 1,3 мкм, коэффициент A_λ равен 5,7 Вт/А и при длине λ, равной 1,55 мкм, коэффициент A_λ равен 4,8 Вт/А. Энергия падающего излучения, соответствующая одному и тому же фототоку, уменьшается с увеличением длины волны.

Мощность шума оптического приемного модуля с pin-фотодетектором и полевым транзистором на входе можно определить с помощью выражения:

$$\overline{i^2}_{pin} = 4 \cdot k \cdot T \cdot (2 \cdot \pi \cdot C_\Sigma) \cdot B^2 \cdot \left(\frac{In_2}{2K} + \frac{In_3 \cdot (2\pi \cdot C_\Sigma) \cdot B \cdot Fn}{Sm} \right), \quad (6)$$

где k– постоянная Больцмана, T – температура; C_Σ – суммарная емкость фотодиода, предварительного усилителя; Fn – шум-фактор полевого транзистора; In_2, In_3 – интегралы Персоника; Sm – крутизна полевого транзистора; K – коэффициент, характеризующий глубину интегрирования во входной цепи фотоприемника.

В результате получаем зависимость чувствительности оптического приемника от скорости передачи:

$$P_{pim} = \frac{A_J}{\eta_m} \sqrt{4 \cdot k \cdot T \cdot (2 \cdot \pi \cdot C_\Sigma) \cdot B^2 \cdot \left(\frac{In_2}{2K} + \frac{In_3 \cdot (2\pi \cdot C_\Sigma) \cdot B \cdot Fn}{Sm} \right)} \quad (7)$$

На рис. 2 приведена, рассчитанная в пакете Mathcad, кривая расчета чувствительности оптического приемника при следующих параметрах: η_m = 0,8; A_λ = 4,8 Вт/А; C_Σ = 0,5 пФ (кривая 1), C_Σ = 1 пФ (кривая 2); In_2 = 0,55, In_3 = 0,085; $Sm = 35 \cdot 10^{-3}$ См; Fn = 1,5.

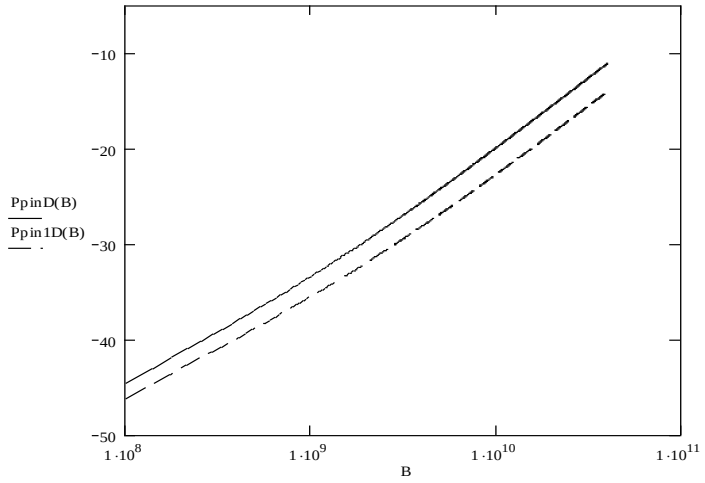


Рис. 2. Зависимость чувствительности оптического приемника с pin-фотодиодом от скорости передачи

Из рис. 2 следует, что чувствительность оптического приемника с увеличением скорости передачи информации быстро уменьшается, что приводит к уменьшению бюджета системы, который равен разности уровней передающего оптического модуля и чувствительности оптического премного устройства.

Расчет чувствительности оптического приемного модуля с оптическим предусилителем можно выполнить, если учесть, что мощность шума i^2 pin-фотодиода уменьшается в G^2 раз (где G – коэффициент усиления оптического усилителя), а шумы усилителя – с учетом эффективности преобразования pin-фотодетектором.

Для тех же параметров , что и в предыдущих расчетах, с учетом того, что полоса спектра сигнала одного канала Δν для системы со скоростью *B* Мбит/с примерно равна Δν = 2 В, построена зависимость чувствительности оптического приемника с pin-фотодиодом от скорости передачи при k_1 = 0,5 (рис. 3).

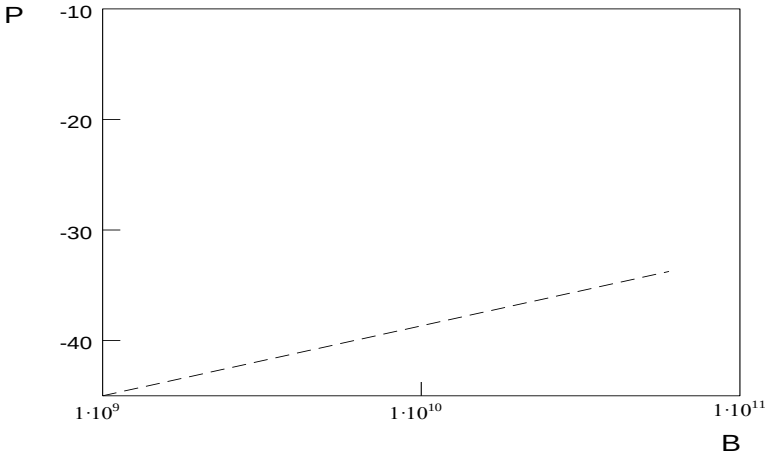


Рис. 3. Зависимость чувствительности оптического приемника с pin-фотодиодом от скорости передачи при k_1 = 0,5

Оценка возможности несанкционированного доступа к информации, передаваемой от абонента, при использовании отраженного от ответвителя сигнала на другой розетке

На основе результатов сравнения оптических приемников различного типа установлено, что приемник с оптическим предусилителем с типичным коэффициентом усиления 20 дБ обеспечивает выигрыш в чувствительности примерно на 7 дБ по сравнению с приемником, использующим лавинный фотодиод и 15 дБ с pin-фотодиодом.

Реальная чувствительность для скорости 2,5 Гбит/с составила – 49 дБм, что обуславливает возможность несанкционированного доступа к восходящему потоку с абонентской розетки. Однако качество приема сигналов будет с большей вероятностью ошибки (~ 10^{-6}).

Заключение

Показано, что в стандартной PON-сети возможен скрытый доступ к передаваемой информации при атаке с любой соседней оптической розетки путем выделения прямого и отраженного сигнала в оптическом сплиттере. Обратного канала, достаточно, чтобы величина этого сигнала была теоретически в рамках квантового предела детектируемости.

ESTIMATION OF THE POSSIBILITY OF UNAUTHORIZED ACCESS TO THE PASSIVE OPTICAL NETWORK

N.N. SERGEEV, V.N. URYADOV

Abstract. The possibility of unauthorized access to information transmitted from the subscriber when using the reflected signal from the coupler to another subscriber outlet is considered. Conduction of the reflected signal level and threshold sensitivity of the receiver is conducted. An assessment of such attack possibility is made.

Keywords: passive optical network, information protection, unauthorized access, network attack, threat.

Список литературы

1. Птицын Г.А. Живучесть динамических сетей телекоммуникаций М.: МТУСИ. 2008.
2. Gutierrez D., Cho J., Kozovsky L.G. // Optical Fiber Communication and the National Fiber Optic Engineers Conference. USA, 25–29 March 2007.
3. Булавкин И.А. // Технологии и средства связи 2006. IW2. С. 104–108.
4. Рекомендация МСЭ-Т G.983.1. Широкополосные оптические сети доступа на базе пассивных оптических сетей.
5. Урядов В.Н., Алишев Я.В., Перспективные информационные технологии в волоконно-оптических сетях телекоммуникаций. Минск, 2003.

СВЕДЕНИЯ ОБ АВТОРАХ

1.	Алисеенко Маргарита Александровна	– магистрант кафедры инфокоммуникационных технологий БГУИР
2.	Аль-Махдави Мустафа Сабах Халил	– аспирант кафедры защиты информации БГУИР
3.	Альмияхи Осама Мажид Хилал	– к.т.н., доцент, институт изобразительных искусств, отдел образования г. Эд Дивания.
4.	Аль-Субаи Амжед Карим Туама	– магистрант кафедры инфокоммуникационных технологий БГУИР
5.	Аль-Фурайджи Одай Джасим Мохаммед	– к.т.н., заместитель заведующего кафедрой информационных технологий университета Шат Аль-Араб.
6.	Аль-Эзайрджави Атир Абдулзахра Салах	– магистрант кафедры инфокоммуникационных технологий БГУИР
7.	Белан Владислав Анатольевич	– магистрант кафедры инфокоммуникационных технологий БГУИР
8.	Белоусова Елена Сергеевна	– к.т.н., доцент кафедры защиты информации БГУИР
9.	Бобов Михаил Никитич	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
10.	Бойправ Владимир Андреевич	– аспирант кафедры защиты информации БГУИР
11.	Вильчиков Александр Иванович	– начальник отдела оптико-механических разработок ОАО «Конструкторское бюро «Дисплей»
12.	Вишняков Владимир Анатольевич	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
13.	Гашков Сергей Валентинович	– начальник отдела разработки авиационных дисплеев и микродисплейных систем индикации ОАО «Конструкторское бюро «Дисплей»
14.	Давронова Лола Ухтамовна	– ассистент Ташкентского университета информационных технологий имени Мухаммада Аль-Хоразмий
15.	Давыдова Надежда Сергеевна	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
16.	Жэнь Сюньхуань	– аспирант кафедры инфокоммуникационных технологий БГУИР
17.	Искрик Андрей Николаевич	– аспирант кафедры инфокоммуникационных технологий БГУИР
18.	Исломов Шахбоз Зокир угли	– аспирант Ташкентского университета информационных технологий имени Мухаммада Аль-Хоразмий
19.	Ковшик Виктория Анатольевна	– аспирант кафедры инфокоммуникационных технологий БГУИР

20.	Конопелько Валерий Константинович	– д.т.н., профессор кафедры инфокоммуникационных технологий БГУИР
21.	Королев Алексей Иванович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
22.	Курилович Андрей Владимирович	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
23.	Липницкий Валерий Антонович	– д.т.н., профессор, заведующий кафедрой высший математики ВА РБ
24.	Ма Цзюнь	– аспирант кафедры инфокоммуникационных технологий БГУИР
25.	Макейчик Екатерина Геннадьевна	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
26.	Мардиев Улугбек Расулович	– старший преподаватель Ташкентского университета информационных технологий имени Мухаммада Аль-Хоразмий
27.	Митюхин Анатолий Иванович	– доцент кафедры физико-математических дисциплин ИИТ БГУИР
28.	Нгуен Ань Туан	– аспирант кафедры инфокоммуникационных технологий БГУИР
29.	Нгуен Тхи Хонг Ньунг	– аспирант факультета радиоэлектроники Технический университет имени Ле Куй Дона.
30.	Нгуен Чунг Тхань	– к.т.н., факультет радиоэлектроники Технический университет имени Ле Куй Дона.
31.	Нгуен Хонг Куан	– магистрант кафедры инфокоммуникационных технологий БГУИР
32.	Никульшин Борис Викторович	– к.т.н., доцент, заведующий кафедрой электронных вычислительных машин БГУИР
33.	Пригон Анна Николаевна	– магистрант кафедры инфокоммуникационных технологий БГУИР
34.	Рабцевич Виолетта Викторовна	– аспирант кафедры инфокоммуникационных технологий БГУИР
35.	Садик Бакир Джафар Садик	– стажер кафедры инфокоммуникационных технологий БГУИР
36.	Саломатин Сергей Борисович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
37.	Самаров Хуснутдин Камардинович	– к.т.н., доцент, Ташкентский университет информационных технологий имени Мухаммадааль-Хоразмий
38.	Сергеев Николай Николаевич	– аспирант кафедры инфокоммуникационных технологий БГУИР
39.	Середа Елена Владимировна	– аспирант кафедры защиты информации БГУИР
40.	Тимофеев Александр Михайлович	– к.т.н., доцент кафедры защиты информации БГУИР
41.	Туча Дмитрий Юрьевич	– аспирант кафедры инфокоммуникационных технологий БГУИР

42.	Урядов Владимир Николаевич	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
43.	Утин Лоенид Львович	– к.т.н., доцент, начальник кафедры связи БГУИР
44.	Фам Хак Хоан	– к.т.н., доцент факультета радиоэлектроники Технического университета имени Ле Куй Дона
45.	Хоменок Михаил Юлианович	– к.т.н., доцент кафедры инфокоммуникационных технологий БГУИР
46.	Хоминич Александр Леонидович	– старший преподаватель кафедры инфокоммуникационных технологий БГУИР
47.	Худойкулов Зариф Туракулович	– к.т.н., доцент, заведующий кафедрой, Ташкентский университет информационных технологий имени Мухаммада Аль-Хоразмий
48.	Цветков Виктор Юрьевич	– д.т.н., заведующий кафедрой инфокоммуникационных технологий БГУИР
49	Шаблюк Анастасия Викторовна	– инженер института тепло- и массообмена имени А.В. Лыкова
50	Шакир Мухаммед Музахем Шакир	– магистрант кафедры инфокоммуникационных технологий БГУИР
51	Эль Масри Абдель Хуссейн Али	– аспирант кафедры инфокоммуникационных технологий БГУИР
52	Эль Хаджи Слейман Кхалед Кхалиль	– аспирант кафедры инфокоммуникационных технологий БГУИР