

DATA COMPRESSION IN BLACK-GRAY-WHITE BARCODING

Dychka I. – Dr. Sc., Professor, Head of the Faculty of Applied Mathematics, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Onai M. – PhD, Associate Professor in the Computer Systems Software Department, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

Sulema OI. – Post-graduate student of the Department of Computer Systems Software, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

ABSTRACT

Context. In this paper the authors propose a method for data compression to be used for presenting information in the form of 2D matrix barcode. The proposed method is based on both a structural-logical approach and using three colors in a barcode instead of two colors as it is in standard black-and-white barcodes. This approach allows to increase data density keeping the same area as bi-color barcodes take. In the paper authors present the data compression method and demonstrate the barcoding technology.

Objective. The goal of the work is to develop a method of data barcoding that would allow to encode more information in the form of 2D matrix barcode.

Method. The method of tricolor matrix barcoding with compression is proposed. The main idea of the method is to compress input textual information on the stage of alphanumeric sequence transformation into a set of barcode patterns, which will form a resulting barcode symbol. It is possible due to intermediate transformation of input characters from initial notation, which is determined by cardinality of an input alphabet, to a notation defined by cardinality of barcode patterns alphabet. Choice of the input alphabet influences overall compression, and it is an important step of the method to choose the initial alphabets for the textual information to be encoded. Use of three colors over standard two colors is also an important component for creating a barcode symbol with increased informational density. As ternary notation is used, the second transformation from the intermediate notation to the ternary one provides more compression. The proposed method allows to represent more textual data in a single barcode symbol than bicolor barcoding approaches do.

Results. The method of tricolor matrix barcoding with compression has been developed and described. Authors provided an example of the method implementation on test data that had been barcoded using the method.

Conclusions. The experiments conducted for this research have confirmed that the proposed method provides more informational density as compared with black-and-white matrix barcodes. The prospects for further research might include studying noise immunity issue in order to guarantee error-free scanning and increased reliability of the barcode, and extending the barcoding software to be used in any alphabet.

KEYWORDS: textual Data Protection, Automated Data Capture, Barcoding, 2D Barcode, Tricolor Barcode, Grayscale Barcode, Black-Gray-White Barcode.

ABBREVIATIONS

BGW-Code is a black-gray-white barcode;

DNM is a decimal numbers mode;

HNM is a hexadecimal numbers mode;

TDM is a textual data mode;

ASM is an ASCII symbols mode.

w_i is a subsequence consisted of elements t_i ;

ω_i is a barcode pattern;

u_z is a subsequence of barcode patterns;

$U_{P_{\Omega_{\text{inf}}}}^{(s)}(P_A)$ is a compression coefficient.

NOMENCLATURE

B is an area of barcode symbol;

s is a number of cells in a barcode pattern;

q is a number of colors, or number base;

V is a maximum capacity of a barcode symbol;

Ω is a symbolism (an alphabet) of barcode;

Ω_{inf} is an alphabet of informational (textual) symbols;

Ω_{aux} is an alphabet of auxiliary (technical) symbols;

P_X is a cardinality of any alphabet X ;

T is an input alphanumeric sequence;

U is a resulting sequence of barcode patterns;

n is a number of adjacent symbols of the same type in the input sequence T ;

m is a number of barcode patterns;

t_i is an element (one character) of the input sequence T ;

INTRODUCTION

One of the important problems in the field of modern information technologies is an issue of information support of automated relocation of objects (goods, freights, medical supplies, documents, parcels etc.) [1]. Automated objects relocation systems are based on using automated identification that enables entering required data into a computer system by its automated scanning from the object.

High demand for automated identification is determined by the desire to improve control on objects relocation, reduce production costs, and increase its profitability and efficiency.

Among various types of automated identification, it is the barcoding technology [2, 3] that has been widely distributed. It happens due to the low cost of both barcode patterns production and scanning equipment.

Barcoding is the way to represent and store information on a carrier using elementary discrete graphical shapes, such as circle, ellipse, square, rectangle, hatch (straight, oblique), triangle, polygon (hexagon, octagon) etc. Information is represented in the form of combinations of elements with different coloring.

Barcoding provides an optical way of information scanning, including distant scanning [4, 5]. A barcode is placed on an object surface, and it is moving along with the object throughout its trajectory.

Since the invention of barcoding (the first patent for the barcode was received in 1952), it has been more than 60 years, however barcoding is still considered as one of the advanced technologies. Moreover, a lot of experts believe that barcodes are among the most prominent discoveries of the 20th century.

There is a number of barcode types which can be divided into 3 main groups: linear, stack, and matrix. A few up to several dozens of alphanumeric codes can be represented in the form of linear barcode. Several hundred characters can be represented in the form of stack barcode. Up to several thousand symbols can be represented in the form of matrix barcode. The subject of this research is matrix barcoding.

Matrix barcode is a two-dimensional array of discrete graphic items combined as one image. The structure of such an array is called a barcode pattern (BC-pattern). The majority of matrix barcodes are black-and-white. However, in recent years there has been growing interest in the development of multicolor matrix barcodes. The most well-known among them are Microsoft's High Capacity Color Barcode [6] and High Capacity Color QR code [7].

Multicolor barcodes provide larger data density in comparison with black-and-white equivalents; however, they have a disadvantage, which is narrow scope of application. This is due to as yet higher cost of color printers compared to black-and-white printers, as well as the high cost of consumables for color printing. Therefore, multicolor barcodes are unable to replace black-and-white barcodes in all areas of applications, considering that black-and-white printing can be more efficient than a color one in certain use cases.

Since modern printing equipment, namely laser printers, provide a high black-and-white printing quality with the required resolution, we consider that it is appropriate to add one more color, the gray, representation of which is not a challenge using black-and-white printer. Thus, we propose to create black-gray-white barcode patterns (Fig. 1) with existing black-and-white printing equipment. As a result, it is expected that matrix barcode data density will be increased.

Subsequently, let us call such black-gray-white barcodes a BGW-Code.

However, to achieve high rate of data density, it is essential to apply a structural-logical approach for increasing barcodes density, in addition to the use of the third color.

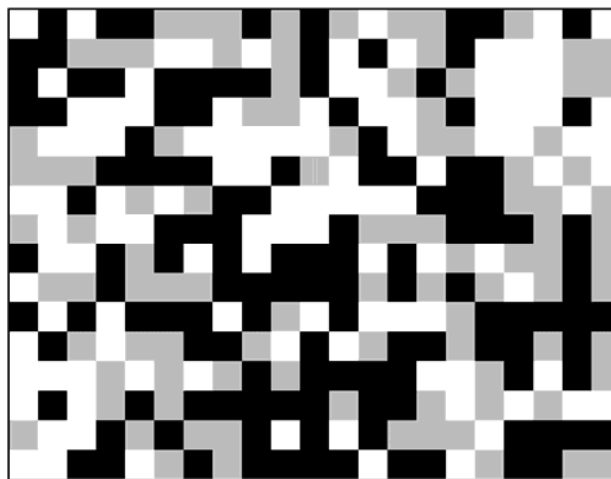


Figure 1 – An example of BGW-Code symbol

The object of study is the process of transforming textual information into a barcode.

The process of a barcode symbol creation becomes more complex because of procedure of data compressing, which is an important part of the method proposed in the paper. Therefore, the process of data barcoding consists of two stages: compression of input information, and transforming compressed data into a barcode symbol.

The subject of study is the methods for data compression and barcode construction.

The methods being developed by authors are aimed at providing a possibility to encode more data into a single barcode symbol.

The purpose of the work is to develop the method of forming barcode symbol with increased informational density.

1 PROBLEM STATEMENT

Requirements to a barcode as a way to store and input information are as follows:

1. Miniaturization of barcode pattern (limited area B is allocated for a barcode pattern ω_i superimposing on an item).
2. Significantly increasing capacity V of a barcode pattern ω_i without changing its geometrical dimensions.
3. Widening the concept of a barcode pattern in order to obtain portable data file (the barcode pattern has to contain not only an access key to information but complete information about the item).

Let us consider a problem of increasing data density of a matrix barcode. Factors of increasing data density could be both a number of colors q used in barcode and the use of specific methods for data compression, which enable increasing capacity V of a barcode pattern ω_i .

In this research we consider tricolor barcode, i.e. $q = 3$, in which black, white, and gray colors are used for elements representation. Such tricolor barcode patterns are easily produced by using an ordinary printer.

Increasing data density basically means increasing of a ratio between an initial data sequence needed to be encoded and a resulting sequence of barcode patterns, which represents by a compression coefficient U .

Therefore, the formal definition of the task to be solved in this research is as follows:

$$\begin{cases} B(\omega_i) \rightarrow \min \\ V(\omega_i) \rightarrow \max, \forall i, \omega_i \in \Omega_{\text{inf}} . \\ U_{P_{\Omega_{\text{inf}}}}^{(s)}(P_A) \rightarrow \max \end{cases}$$

Thus, in this paper we present the new method of alphanumeric data compression; these data are a subject of representation in the form of black-gray-white barcode.

2 REVIEW OF THE LITERATURE

Methods of information barcoding as well as barcodes themselves are the subject of research for many scientists.

In [8], the author presents a method to generate and decode two-dimensional color barcode consisted of several blocks, which are a black-and-white configuration block that encodes auxiliary information about the barcode itself and a set of color data blocks that encodes actual data.

In the patent [9] it is proposed to store information decoded from a barcode in a form of character-based data in an auxiliary field (e.g. a comment field).

The authors of [10] propose a new approach of decoding color barcode, which does not require a reference color palette. They describe an algorithm, in which groups of color bars are decoded at once, what is exploiting the fact that joint color changes can be represented by a low-dimensional space.

A prototype for generating and reading a HCC2D code format on both PC and mobile phones is presented in [11]. The authors provide experimental results considering different operating scenarios and data densities in comparison with 2-dimensional barcodes.

The authors of [12] describe a method of high capacity color barcodes generation, which operates due to embedding independent data into two different printer colorant channels via halftone-dot orientation modulation.

In [13], an approach for localization and segmentation of a 2D color barcode when it is read using computer vision techniques is presented. The authors develop a progressive strategy to achieve high accuracy in diverse scenarios and computational efficiency.

The authors in [14] propose both a system and a method to encode and decode data in a color barcode pattern using dot orientation and color separability. They aver the method to be robust against interseparation misregistration with a small symbol error rate.

COBRA system, which is a visible light communication (VLC) system for off-the-shelf smartphones, is presented in [15]. The proposed system is able to encode data into specially designed 2D color barcodes. To achieve it, the authors developed a new COBRA barcode optimized for streaming between small-size screen and low-speed camera of smartphones.

As presented, a lot of various solutions for barcoding exist, however there still are different relevant problems concerning data barcode representation improvement, which requires some new approaches in data compressing.

3 MATERIALS AND METHODS

A barcode symbol consists of barcode patterns. In its turn, a barcode pattern consists of s elements, which are matrix cells on a carrier. Each cell can be either black, gray, or white.

We assume that maximum capacity of a barcode symbol equals V barcode patterns. In this case, $V \leq 3^s$, where 3 is a number of colors and s is a number of cells in the barcode pattern (Fig. 2). As shown in Table 1, $V_{\text{max}} = 3^s$ barcode patterns. A set of all possible barcode patterns with a fixed s forms the multiplicity, or the alphabet Ω of cardinality $P_{\Omega} = 3^s$. Let us call this alphabet a symbolism of barcode.

The symbolism of barcode consists of informational patterns Ω_{inf} and auxiliary patterns Ω_{aux} , i.e. $\Omega = \Omega_{\text{inf}} \cup \Omega_{\text{aux}}$. Capacity of informational patterns is $P_{\Omega_{\text{inf}}}$ and capacity of auxiliary patterns is $P_{\Omega_{\text{aux}}}$. Thus, $P_{\Omega_{\text{inf}}} + P_{\Omega_{\text{aux}}} = 3^s$.

To represent information on a carrier, we use $P_{\Omega_{\text{inf}}}$ informational barcode patterns.

Auxiliary patterns are used to switch between encoding modes, indicate START and STOP barcode patterns and setup a scanner.

Table 1 – Dependence of barcode pattern maximal capacity on barcode pattern digital capacity

s value	3^s value	Maximal capacity of a barcode pattern, V_{max}	Type of BGW-Code
4	3^4	81	Very small
5	3^5	243	Small
6	3^6	729	Middle
7	3^7	2187	Large
8	3^8	6561	Very large
9	3^9	19683	Ultra large

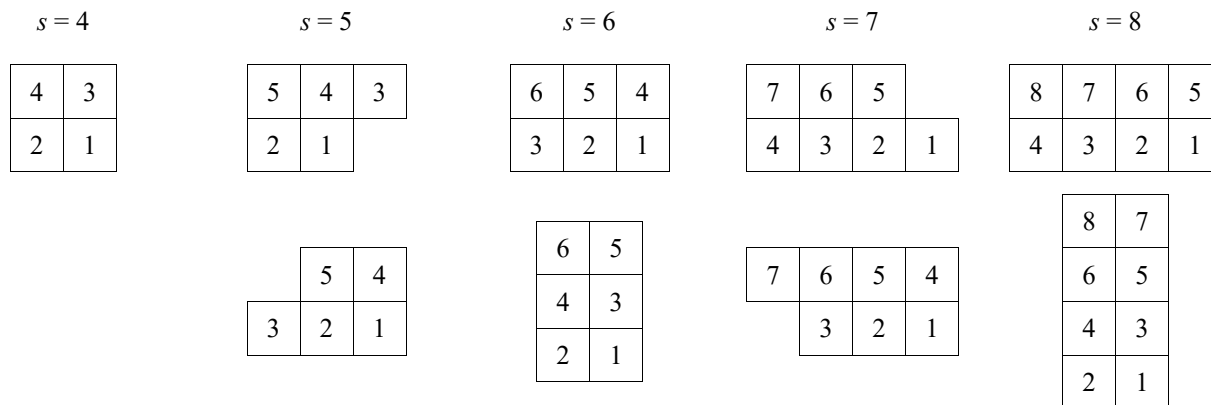


Figure 2 – Bits arrangement in a barcode pattern according to digit capacity s

Let us represent an input h long alphanumeric sequence $T = t_1 t_2 \dots t_h$. Elements of the sequence belong to extended ASCII, i.e. $t_i \in \text{ASCII}(256)$, $i = 1, 2, \dots, h$.

Let the set ASCII be presented as the following: $\text{ASCII} = \{L \cup D \cup C\}$ where L is a letters set, D is a digits set, and C is a special symbols set.

The sequence T divides into adjacent subsets that are consisted of elements belonged to one of ASCII subsets:

$$T = w_1 w_2 \dots w_k, \quad (1)$$

where $w_i = t_1 t_2 \dots t_n$ is a subsequence of the input sequence which contains elements t_i of only one set of ASCII subsets, namely L , D or C . The subsequences w_1, w_2, \dots, w_k can be situated in any order in the input sequence T .

Let alphanumeric symbols t_i belong to an alphabet A ,

which belongs to ASCII, i.e. $t_i \in A$, $A \subset \text{ASCII}$. Cardinality of the alphabet A is considered to be equal P_A . The alphabet A corresponds with a numeric set $\{0, 1, \dots, P_A - 1\}$ that represents numbers of the symbols of the alphabet A in the alphabet.

Now we turn the subsequence $w_i = t_1 t_2 \dots t_n$ formed out of the symbols of the alphabet A into a barcode. In the barcode form, the subsequence $t_1 t_2 \dots t_n$ corresponds to a subsequence u_z consisted of m barcode patterns: $u_z = \omega_1 \omega_2 \dots \omega_m$ where $\omega_i \in \Omega_{\text{inf}}$.

In turn, the alphabet Ω_{inf} corresponds with a numeric set $\{0, 1, \dots, P_{\Omega_{\text{inf}}} - 1\}$, as barcode patterns of the barcode symbolism, which are used for representing textual data, can be numbered from 0 to $P_{\Omega_{\text{inf}}}$.

The above-mentioned transformation is considered as $w_i \rightarrow u_z$, i.e. $(t_1 t_2 \dots t_n) \rightarrow (\omega_1 \omega_2 \dots \omega_m)$. Then the con-

dition for transforming n adjacent symbols belonging to the alphabet A into m barcode patterns belonging to the alphabet Ω_{inf} (i.e. the barcode symbolism) is as follows:

$$n(P_A) \rightarrow m(P_{\Omega_{\text{inf}}}). \quad (2)$$

Practically, the transformation (2) means a transformation of n -digits number in a notation P_A into m -digits number in a notation $P_{\Omega_{\text{inf}}}$.

The transformation (2) will be with compression if $n \log_3 P_A \lceil > m \rceil \log_3 P_{\Omega_{\text{inf}}}$ and at the same time,

$P_A^n - 1 \leq P_{\Omega_{\text{inf}}}^m - 1$ where $P_A^n - 1$ and $P_{\Omega_{\text{inf}}}^m - 1$ are quantitative equivalents of, correspondingly, maximal n -digits number in a notation P_A and maximal m -digits number in a notation $P_{\Omega_{\text{inf}}}$. $n \log_3 P_A \lceil$ is a length of the ternary sequence, which corresponds to an alphanumeric sequence $w_i = t_1 t_2 \dots t_n$.

To let n symbols long alphanumeric subsequence $w_i = t_1 t_2 \dots t_n$ be transformed into m long barcode patterns subsequence $u_z = \omega_1 \omega_2 \dots \omega_m$ with compression, it is necessary that the following condition is met:

$$\begin{cases} P_A^n - 1 \leq P_{\Omega_{\text{inf}}}^m - 1, \\ n \log_3 P_A \lceil > m \rceil \log_3 P_{\Omega_{\text{inf}}} \lceil. \end{cases}$$

It is important that $n \log_3 P_A \lceil = s$. Thus, it is necessary and sufficient condition for transforming the subsequence $w_i = t_1 t_2 \dots t_n$ of alphanumeric symbols from the alphabet A into the subsequence $\omega_1 \omega_2 \dots \omega_m$ of barcode patterns from the symbolism Ω_{inf} with compression:

$$\begin{cases} P_A^n \leq P_{\Omega_{\text{inf}}}^m; \\ n \log_3 P_A \lceil > ms, \end{cases} \quad (3)$$

where ms is a number of tricolor cells on a carrier that represent the subsequence w_i .

Such a transformation is necessary for ensuring space-keeping data representation on a carrier and increasing data density of barcode patterns with their fixed geometrical dimensions and at the same time, unchanging carrier size.

We define a degree of input data compression as a ratio of a length of ternary sequence that corresponds with alphanumeric sequence w_i to a number of cells on a carrier that represents subsequence w_i in barcoded form and refer to it as a compression coefficient:

$$U_{P_{\Omega_{\text{inf}}}}^{(s)}(P_A) = \frac{n \lceil \log_3 P_A \rceil}{ms} \quad (4)$$

Thus, the data compression problem (3) consists in finding such P_A with fixed $P_{\Omega_{\text{inf}}}$, as well as parameters n and m , so that maximum value of $U_{P_{\Omega_{\text{inf}}}}^{(s)}(P_A)$ is guaranteed.

4 EXPERIMENTS

Let us consider $s = 7$, i.e. large BGW-Code will be considered, thereby achieving barcode symbols with capacity of 2187 barcode patterns (see Table 1). Thus, the symbolism of the barcode comprises 2187 tricolor barcode patterns, each of which consists of 7 cells. The barcode patterns correspond to a numeric set $\{0, 1, \dots, 2186\}$.

15 barcode patterns let be considered as auxiliary ones. In this case, 2172 barcode patterns remain for representing information.

The inequality system (3) for large BGW-Code is as follows:

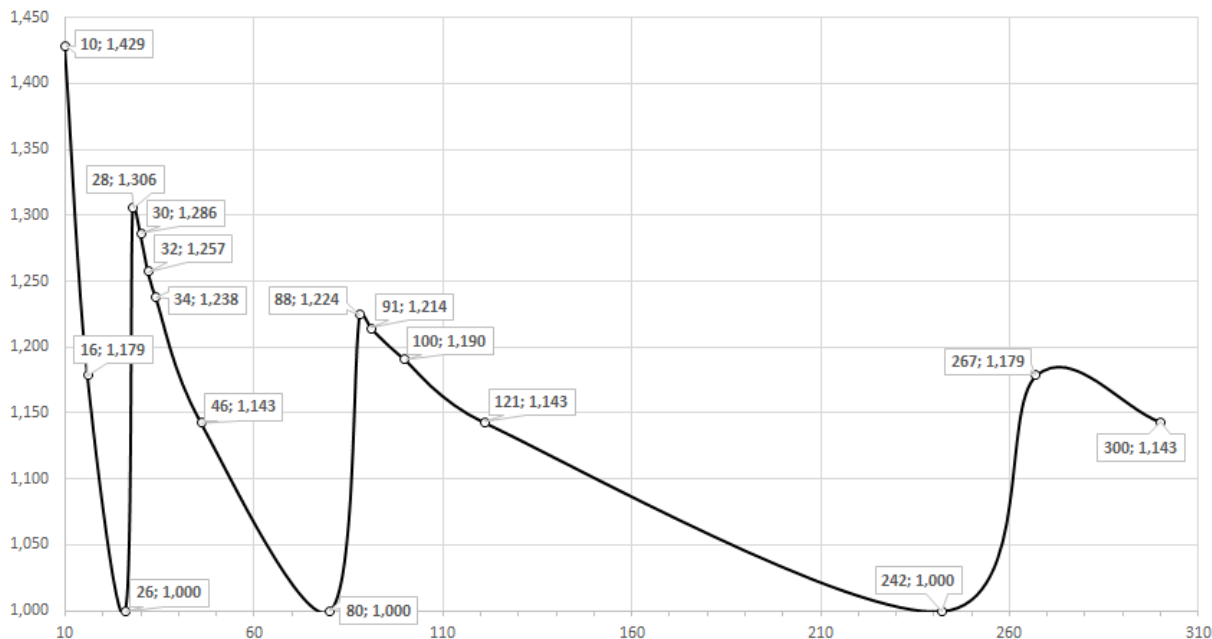


Figure 3 – Dependence of a compression coefficient on a cardinality P_A of the alphabet when $s = 7$

$$\begin{cases} P_A^n \leq 2172^m \\ \lfloor n \rfloor \log_3 P_A \lceil > 7m \end{cases} \quad (5)$$

Now we need to solve the system (5) relatively to P_A . Only integer values are considered as solutions, and the compression coefficient is calculated according to (4):

$U_{2172}^{(s)}(P_A) = \frac{n \lceil \log_3 P_A \rceil}{7m}$. Such n and m values are sought for each P_A , so as maximum compression coefficient will be achieved. Fig. 3 defines dependence of a compression coefficient on a cardinality P_A .

The Table 2 below shows some integer solutions (P_A, n, m) that provide compression of data when it is represented as a barcode.

If w_i is the subsequence of decimal numbers, maximum compression with a compression coefficient 1.429 will be achieved when each 10-digits subsequence of adjacent decimal numbers corresponds with 3 barcode patterns. Practically, transformation “10” → “3” means that 10-digits decimal number transforms into 3-digits number in a notation 2172.

The alphabet $P_A = 28$ that is composed of 16 hexadecimal numbers $\{0 - F\}$ and 12 other random symbols, such as letters and special characters, provides a possibility to represent hexadecimal sequences, e.g. exe-files, in the form of barcode. In this regard, we should use the transformation “16” → “7” where each 16-digits subsequence corresponds with 7 barcode patterns, which means that 16-digits hexadecimal number is transformed into 7-digits number in a notation 2172. With this transformation, 16-digits subsequence is compressed with a compression coefficient 1.306 (see Table 2).

The alphabet $P_A = 267$ provides a possibility to represent an input information comprised of any ASCII(256) symbols in the form of barcode.

In this case, the compression coefficient equals 1.179 and each subsequence consisted of 11 alphanumeric symbols corresponds with 8 barcode patterns, i.e. the transformation “11” → “8” means that 11-digits number in the notation 267 transforms into 8-digits number in the notation 2172.

From Fig. 3 and Table 2, we can assume that it is most appropriate to use the following 4 compression modes for an input alphanumeric sequence T (see Fig. 4):

- DNM, Decimal Numbers Mode: $P_A = 10$, the mode for compressing decimal subsequences,
- HNM, Hexadecimal Numbers Mode: $P_A = 28$, the mode for compressing hexadecimal subsequences,
- TDM, Textual Data Mode: $P_A = 88$, the mode for compressing alphanumeric data (overall number of symbols in the text shall not exceed 88),
- ASM, ASCII Symbols Mode: $P_A = 267$, the mode for compressing subsequences composed of ASCII(256) symbols.

To switch between the modes, switch symbols $\omega_{2172} - \omega_{2181}$ that corresponds to appropriate auxiliary barcode patterns, the mode switchers, are used. For instance, the mode switcher ω_{2172} corresponds to the barcode pattern number 2172 in the barcode symbolism and provides a transition from ASM to DNM.

Before a barcode image be plotting on a carrier, the input alphanumeric sequence (1), which shall be represented as a barcode, is reduced to the following form:

$\tilde{T} = \omega w_1 \omega w_2 \dots \omega w_k$, where ω is a mode switcher, $\omega \in \{\omega_{2172}, \dots, \omega_{2181}\}$. Each subsequence $w_i = t_1 t_2 \dots t_n$ is encoded under the rules of the appropriate mode.

In DNM, the following transformation is performed:

$$\sum_{i=1}^{10} t_i 10^{i-1} \rightarrow \sum_{r=1}^3 \omega_r 2172^r \text{ where } t_i \in \{0, 1, 2, \dots, 9\}.$$

In HNM, the transformation

$$\sum_{i=1}^{16} t_i 28^{i-1} \rightarrow \sum_{r=1}^7 \omega_r 2172^r \text{ is performed, where } t_i \in \{0, 1, 2, \dots, 27\}.$$

In TDM, the transformation $\sum_{i=1}^{12} t_i 88^{i-1} \rightarrow \sum_{r=1}^7 \omega_r 2172^r$

is performed, where $t_i \in \{0, 1, 2, \dots, 87\}$.

In ASM, the following transformation is performed:

$$\sum_{i=1}^{11} t_i 267^{i-1} \rightarrow \sum_{r=1}^8 \omega_r 2172^r \text{ where } t_i \in \{0, 1, 2, \dots, 266\}.$$

As a result of these transformations, the array of number from the range $0 \div 2171$ is obtained instead of the input alphanumeric sequence T . Then, each number is replaced by the appropriate barcode pattern from the symbolism and is arranged on a carrier. The barcode symbol can have either square or rectangle shape.

Table 2 – Some integer solutions of the system (5) for $s = 7$

Cardinality of alphabet A , P_A	Type of transformation, $n \rightarrow m$	Compression coefficient, $U_{2172}^{(7)}(P_A)$	What can be represented
10	10 → 3	1.429	Decimal numeric sequences
28	16 → 7	1.306	Hexadecimal numeric sequences
30	9 → 4	1.286	Textual information comprised of uppercase Latin-script letters
32	11 → 5	1.257	
34	13 → 6	1.238	
88	12 → 7	1.224	Textual information comprised of Latin alphabet and/or Cyrillic alphabet
91	17 → 10	1.214	
100	5 → 3	1.190	Shortened ASCII(128)
267	11 → 8	1.179	ASCII(256)

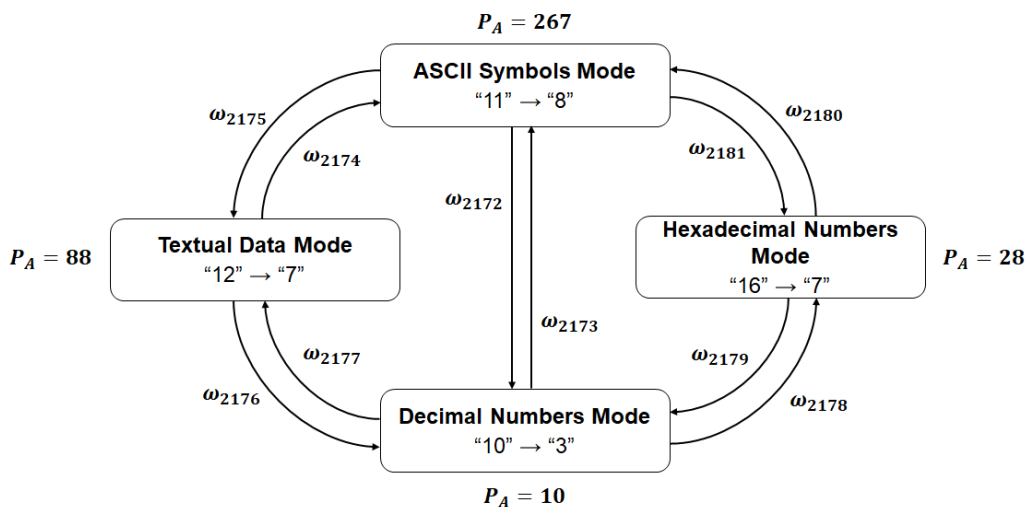


Figure 4 – Interconnection between the compression modes

The method of compression of alphanumeric information, which is to be represented as a BGW-Code, is as follows:

1. Taking into account the parameter V , which is the capacity of a barcode symbol, the cardinality of the symbolism Ω is defined as $P_{\Omega} = 3^s$ where $s = \lceil \log_3 V \rceil$.

2. A necessary number $P_{\Omega_{\text{aux}}}$ of auxiliary barcode patterns shall be chosen. A number of informational barcode patterns is $P_{\Omega_{\text{inf}}} = P_{\Omega} - P_{\Omega_{\text{aux}}}$.

3. Local extrema of a compression coefficient $U_{P_{\Omega_{\text{inf}}}}^{(s)}$ are found. For this purpose, the system (3) is solved with established s and $P_{\Omega_{\text{inf}}}$. Solutions are a set of alphabet cardinality P_{A_i} values, under which a local extremum of a compression coefficient $U_{P_{\Omega_{\text{inf}}}}^{(s)}$ is achieved, and a type of transformation “ n ” \rightarrow “ m ” for each P_{A_i} .

4. A number ψ of the compression modes is determined, where ψ is a number of chosen local extrema of the compression coefficient $U_{P_{\Omega_{\text{inf}}}}^{(s)}$. Modes transmission rules shall be defined.

5. An appropriate alphabet is formed for each mode.

6. Rules of partitioning input alphanumeric sequences to subsequences of adjacent symbols, which consists only of symbols of the appropriate mode alphabet, are formed.

7. Each obtained subsequence is processed by the rules of the appropriate mode and transformed into numeric form, which is a sequence of numbers from the range $0 \div P_{\Omega_{\text{inf}}} - 1$.

The proposed method of data compression can be used for random symbol sequences input using a keyboard.

5 RESULTS

Let us consider an input textual information consisted of 82 ASCII symbols, which shall be represented in the form of BGW-Code (all the data are fictional):

OLEKSIY KLYMENKO, 01/02/1990,
 XX83629, 36 KHRESHCHATYK STR, (6)
 0156318044 KYIV-21 UKR.

As a result of the analysis of the alphanumeric sequence (6), which is conducted by the appropriate software, the syntax analyzer, the following string is obtained:

$\xrightarrow{12}$ OLEKSIY_KLYMENKO_01/02/1990_XX83629 $\xrightarrow{12}$
 $\xrightarrow{12}$ 9_36_KHRESHCHATYK_STR_ω₂₁₇₆0156318044ω₂₁₇₇KYIV-21_UKR $\xrightarrow{12}$

In this string, 2 mode switchers were inserted: ω_{2176} , to switch from TDM to DNM, and ω_{2177} , to switch from DNM to TDM.

Each 12-digits subsequence (there are 6 such subsequences) shall be replaced by 7 numbers (the transformation “12” \rightarrow “7”) from the range $0 \div 2171$, and each 10-digits subsequence consisted of decimal numbers shall be replaced by 3 numbers (the transformation “10” \rightarrow “3”) from the same range $0 \div 2171$.

Eventually, a numeric sequence comprised of 47 numbers from the range $0 \div 2186$ (as 2172 informational barcode patterns with numbers $0 \div 2171$ and 15 auxiliary barcode patterns with numbers $2171 \div 2186$ compose the symbolism Ω), including 2 mode switchers ω_{2176} and ω_{2177} , which correspond to numbers 2176 and 2177, is obtained as follows:

```

0202221 2111100 1212101 1012020 1221120 0200011
2021120 0110022 2212011 0010200 2100120 0110020
1000111 2102222 0000002 1201022 1100022 1001010
0020101 1010022 2200200 0000211 2221122 1012121
2002011 1202111 2022222 2022212 0101121 1210121
2020021 2220112 0210210 2002012 0122212 2222121
0001020 0101212 1212222 2222122 0012101 0210100
1002002 1212012 1202122 0212202 1100201.
    
```

To obtain a barcode symbol, each of 47 symbols shall correspond to a barcode pattern consisted of 7 tricolor cells (see Fig. 2).

Since a barcode symbol acquires the rectangle shape, one more barcode pattern shall be added to 47 patterns of the barcode: ω_{2182} that represents Pad symbol, a placeholder.

Thus, the barcode symbol presented in Fig. 1 comprised of 336 tricolor cells, as 7×48 barcode patterns is equal to 336. The dimension of the barcode symbol is 16×21 cells.

6 DISCUSSION

Let us consider the obtained results in order to discuss efficiency of the method.

If the textual sequence (6) consisted of 82 symbols of the alphabet with cardinality $P_A = 88$ is represented on a carrier as a black-and-white barcode image, it would require $82 \lceil \log_2 88 \rceil = 574$ black-and-white cells. If the same sequence (6) is represented as a tricolor BGW image, it would require $82 \lceil \log_2 88 \rceil = 410$ tricolor cells. In other words, data density of the barcode symbol increases approximately in $574/410 = 1.4$ times. It happens due to transition from two-color to tricolor image.

As a result of the use of both three colors and the proposed compression method, it takes 336 tricolor cells (see Fig. 1) to represent the textual sequence (6).

Thus, the proposed method provides data density with the compression coefficient $410/336 = 1.22$. The total effect of the transition from two-color to tricolor image alongside using the compression method provides compression with the coefficient $574/336 = 1.708$.

Increasing data density by 1.708 times is assured due to trichromatism (1.4) and the compression method (1.22). Indeed, $1.4 \times 1.22 = 1.708$.

Thus, the multicolor barcoding method proposed in the paper allows to perceptibly increase amounts of information that can be stored in the form of barcode. In the example above, the data density effect is up to 70%. Depending on the area of application and a specific use case, it can ensure significant benefits, such as autonomous access to large amount of actual data, instead of keeping some general information with a link to more data, which is much more convenient, reliable, and in some cases, even more secure way to get information.

CONCLUSIONS

As barcodes are widely used in multiple fields of human activity, there still are various issues concerned with encoding information. And one of such problems is barcoding more data using the same area of barcode graphical representation.

The scientific novelty of obtained results is that the method of tricolor barcoding with compression is firstly proposed. Compression is achieved due to input data transformation into barcode patterns. The proposed transformation method provides transforming a subsequence of input characters into a shorter subsequence of barcode patterns which will form then the resulting barcode symbol. Use of three colors in the barcode ensures additional compression due to use of ternary notation. Combination of these two approaches allows to barcode more information using the same area of a barcode symbol then it would be with use of binary notation without compression.

The practical significance of the proposed method is that more textual information can be encoded in the form of a single barcode symbol. It can be successfully used in various practical applications when size of the overall barcode symbol is essential, especially when there are quite a lot of data to be barcoded.

Prospects for further research are to study noise immunity issue, which must be considered in order to guarantee error-free scanning and increase reliability of the barcode, and to extend the barcoding software to be used in any language, not only Latin and Cyrillic alphabets but also some specific alphabets, such as Korean, Georgian, Arabic etc., and hieroglyphics.

REFERENCES

1. Forrest P. J., Campbell M. J., Fullerton T. J., Celenzano M. J., Brewer R. K. U.S. Patent 6,049,781. Relocation tracking system and method /; applicant HP Enterprise Services LLC. No. US08/634,479 ; appdate 18.04.1996 ; pubdate 11.04.2000.
2. Sriram Th., et al. Applications of barcode technology in automated storage and retrieval systems, *Industrial Electronics, Control, and Instrumentation : The 22nd International Conference, 9 August 1996 : proceedings*. Taipei, IEEE IECON, 1996, Vol. 1.
3. Sun H.-Y. The application of barcode technology in logistics and warehouse management, *Education Technology and Computer Science : The First International Workshop, 7–8 March 2009 : proceedings*. Wuhan, IEEE, 2009, Vol. 3.
4. Kaminsky M. A., Choi J., Lim S., Palmer M. C. U.S. Patent USD710362S1. Barcode scanning device; applicant Motorola Solutions Inc. No. US29/458,380 ; appdate 19.06.2013 ; pubdate 05.08.2014.
5. Bridgelall R., Katz J., Goren D., Dvorkis P., Li Y. U.S. Patent US5988508A. Laser scanning system and scanning method for reading 1-D and 2-D barcode symbols /; applicant Symbol Technologies LLC. No. US08/871,615 ; appdate 10.06.1997 ; pubdate 23.11.1999.
6. Grillo A., Lentini A. et al. High Capacity Colored Two Dimensional Codes, *Computer Science and Information Technology : The International Multiconference, 18–20 October 2010, proceedings*. Wisla, IEEE, 2010, pp. 709–716.
7. High capacity color barcodes [Electronic resource]. Access mode: <http://research.microsoft.com/en-us/projects/hccb/>.
8. Cattrone P. U.S. Patent US7478746B2. Two-dimensional color barcode and method of generating and decoding the same /; applicant Konica Minolta Laboratory USA Inc. No. US11/444,288 ; appdate 31.05.2006 ; pubdate 06.12.2007.
9. Barrus J., Wolff G. J.; U.S. Patent US7150399B2. Embedding barcode data in an auxiliary field of an image file / applicant Ricoh Co Ltd. No. US10/865,584 ; appdate 09.06.2004 ; pubdate 15.12.2005.
10. Bagherinia H., Manduchi R. A theory of color barcodes, *Computer Vision Workshops (ICCV Workshops) : International Conference, 6–13 November 2011 : proceedings*. Barcelona, IEEE, 2011.
11. Querini M., Grillo A. et al. 2D Color Barcodes for Mobile Phones, *International Journal of Computer Science and Applications*, 2011, Vol. 8, No. 1, pp. 135–155.
12. Bulan O., Monga V., Sharma G. High capacity color barcodes using dot orientation and color separability, *Media Forensics and Security : Symposium, 19–21 January 2009 : proceedings*. San Jose, SPIE, 2009, Vol. 7254.
13. Parikh D., Jancke G. Localization and segmentation of a 2D high capacity color barcode, *Applications of Computer Vision : The IEEE Workshop, 7–9 January 2008 : proceedings*. Copper Mountain, IEEE, 2008.
14. Bulan O., Monga V., Sharma G. U.S. Patent US8100330B2. Method for encoding and decoding data in a color barcode pattern; applicant Xerox Corp. No. US12/436,456 ; appdate 06.05.2009 ; pubdate 11.11.2010.
15. Hao T., Zhou R., Xing G. COBRA: Color barcode streaming for smartphone systems, *Mobile systems, applications, and services (MobiSys '12) : The 10th International Conference, 25–29 June 2012 : proceedings*. Low Wood Bay, ACM, 2012, pp. 85–98.

Received 27.06.2019.
Accepted 05.02.2020.

УДК 004.627

УЩІЛЬНЕННЯ ДАНИХ ПРИ ЧОРНО-СИРО-БІЛОМУ ШТРИХОВОМУ КОДУВАННІ ДАНИХ

Дичка І. А. – д-р техн. наук, професор, декан факультету прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Онай М. В. – канд. техн. наук, доцент кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Сулема О. К. – аспірантка кафедри програмного забезпечення комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу отримання триколірних матричних штрихових кодів підвищеної інформаційної щільності для представлення текстової інформації. Запропонований метод базується на структурно-логічному підході та використанні трьох кольорів у штрихкодівій позначці замість двох, які широко використовуються у чорно-білих штрихових кодах. Такий підхід дозволяє збільшити інформаційну щільність, зберігаючи ту саму площу, що мав би чорно-білий штриховий код.

Метод. Запропоновано метод триколірного матричного штрихового кодування з ущільненням. Основна ідея методу полягає у ущільненні вхідної текстової інформації на етапі перетворення алфавітно-цифрової послідовності у сукупність штрихкодівих знаків, які сформують результуючу штрихкодіву позначку. Це є можливим завдяки проміжному перетворенню вхідних символів з початкової системи числення, яка визначається потужністю вхідного алфавіту, у систему числення, визначену потужністю алфавіту штрихкодівих знаків. Від вибору початкового алфавіту залежить остаточне ущільнення, тому обрання алфавітів для текстової інформації, яка кодується, є важливим кроком запропонованого методу. Використання трьох кольорів (чорного, сірого та білого) замість двох (чорного та білого) також відіграє значну роль у процесі створення штрихкодівої позначки з підвищеною інформаційною щільністю. Оскільки використовується трійкова система числення, друге перетворення вхідної послідовності з проміжної системи числення у трійкову систему числення забезпечує ще більше ущільнення. Запропонований метод дозволяє представляти більше текстової інформації у вигляді однієї штрихкодівої позначки, ніж можна представити, використовуючи підходи двоколірного штрихового кодування.

Результати. Розроблено та описано метод триколірного матричного штрихового кодування з ущільненням. Наведено приклад практичного застосування методу на наборі тестових даних із демонстрацією відповідної триколірної матричної штрихкодівої позначки.

Висновки. Проведені в рамках цього дослідження експерименти підтвердили, що запропонований метод забезпечує більшу інформаційну щільність порівняно з чорно-білими матричними штриховими кодами. Перспективи подальших досліджень можуть полягати у вивченні питання завадостійкості, яка необхідна для забезпечення безпомилкового сканування та підвищеної надійності штрихового коду, а також у розширенні програмного забезпечення для штрихового кодування на інші мови та нетипові алфавіти.

КЛЮЧОВІ СЛОВА: захист текстової інформації, автоматичне введення даних, штрихове кодування, матричні штрихові коди, триколірні штрихові коди, штрихові коди у градієнті сірого, чорно-сіро-білі штрихові коди.

УДК 004.627

СЖАТИЕ ДАННЫХ ПРИ ЧЕРНО-СЕРО-БЕЛОМ ШТРИХОВОМ КОДИРОВАНИИ

Дичка И. А. – д-р техн. наук, профессор, декан факультета прикладной математики Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина.

Онай Н. В. – канд. техн. наук, доцент кафедры программного обеспечения компьютерных систем Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина.

Сулема О. К. – аспирант кафедры программного обеспечения компьютерных систем Национального технического университета Украины «Киевский политехнический институт имени Игоря Сикорского», Киев, Украина.

АННОТАЦИЯ

Актуальность. Рассмотрена задача получения трехцветных матричных штриховых кодов с повышенной информационной плотностью для представления текстовой информации. Предложенный метод основывается на структурно-логическом подходе и на использовании трех цветов в штрихкодировании вместо двух, которые широко используются в черно-белых штриховых кодах. Такой подход позволяет увеличить информационную плотность, сохраняя ту же самую площадь, которую занимал бы черно-белый штриховой код.

Метод. Предложен метод трехцветного матричного штрихового кодирования со сжатием. Основная идея метода заключается в сжатии введенной текстовой информации на этапе превращения алфавитно-цифровой последовательности в совокупность штрихкодированных знаков, которые формируют конечное штрихкодированное изображение. Это возможно благодаря промежуточному преобразованию входящих символов из исходной системы исчисления, которая определяется мощностью входящего алфавита, в систему исчисления, определенную мощностью алфавита штрихкодированных знаков. От выбора исходного алфавита зависит окончательное сжатие, поэтому выбор алфавитов для кодируемой текстовой информации является важным шагом предложенного метода. Использование трех цветов (черного, серого и белого) вместо двух (черного и белого) также играет немаловажную роль в процессе формирования штрихкодированного изображения с повышенной информационной плотностью. Поскольку используется троичная система исчисления, второе преобразование входящей последовательности из промежуточной системы исчисления в троичную систему исчисления обеспечивает еще большее сжатие. Предложенный метод позволяет представлять большую текстовую информацию в виде одного штрихкодированного изображения, чем можно представить, используя подходы двухцветного штрихового кодирования.

Результаты. Разработан и описан метод трехцветного матричного штрихового кодирования со сжатием. Приведен пример практического использования метода на наборе тестовых данных с демонстрацией соответствующего трехцветного матричного штрихкодированного изображения.

© Dychka I., Onai N., Sulema O., 2020
DOI 10.15588/1607-3274-2020-1-13

Выводы. Проведенные в рамках этого исследования эксперименты подтвердили, что предложенный метод обеспечивает большую информационную плотность по сравнению с черно-белыми матричными штриховыми кодами. Перспективы дальнейших исследований могут заключаться в изучении вопроса помехоустойчивости, которая необходима для обеспечения безошибочного сканирования и повышенной надежности штрихового кода, а также в расширении программного обеспечения для штрихового кодирования на другие языки и нетипичные алфавиты.

КЛЮЧЕВЫЕ СЛОВА: защита текстовой информации, автоматический сбор данных, штриховое кодирование, матричные штриховые коды, трехцветные штриховые коды, штриховые коды в градиенте серого, черно-серо-белые штриховые коды.

ЛИТЕРАТУРА / LITERATURA

1. Патент США 6,049,781. Relocation tracking system and method / P. J. Forrest, M. J. Campbell, T. J. Fullerton, M. J. Celentano, R. K. Brewer ; заявник HP Enterprise Services LLC. – № US08/634,479 ; заявл. 18.04.1996 ; опубл. 11.04.2000.
2. Applications of barcode technology in automated storage and retrieval systems / [Th. Sriram, et al.] // Industrial Electronics, Control, and Instrumentation : XXII міжнародна конференція, 9 серпня 1996 р. : тези доповідей. – Тайбей : IEEE IECON, 1996. – Т. 1.
3. Sun H.-Y. The application of barcode technology in logistics and warehouse management / H.-Y. Sun // Education Technology and Computer Science : I-й міжнародний семінар, 7–8 березня 2009 р. : тези доповідей. – Ухань : IEEE, 2009. – Т. 3.
4. Патент США USD710362S1. Barcode scanning device / M. A. Kaminsky, J. Choi, S. Lim, M. C. Palmer ; заявник Motorola Solutions Inc. – № US29/458,380 ; заявл. 19.06.2013 ; опубл. 05.08.2014.
5. Патент США US5988508A. Laser scanning system and scanning method for reading 1-D and 2-D barcode symbols / R. Bridgelall, J. Katz, D. Goren, P. Dvorkis, Y. Li ; заявник Symbol Technologies LLC. – Application US08/871,615 ; заявл. 10.06.1997 ; опубл. 23.11.1999.
6. High Capacity Colored Two Dimensional Codes / [A. Grillo, A. Lentini, et al.] // Computer Science and Information Technology : Міжнародна мультиконференція, 18–20 жовтня 2010 р. : тези доповідей. – Вісла : IEEE, 2010. – С. 709–716.
7. High capacity color barcodes [Електрон. ресурс]. – Режим доступу: <http://research.microsoft.com/en-us/projects/hccb/>.
8. Патент США US7478746B2. Two-dimensional color barcode and method of generating and decoding the same / P. Catrone ; заявник Konica Minolta Laboratory USA Inc. – № US11/444,288 ; заявл. 31.05.2006 ; опубл. 06.12.2007.
9. Патент США US7150399B2. Embedding barcode data in an auxiliary field of an image file / J. Barrus, G. J. Wolff ; заявник Ricoh Co Ltd. – № US10/865,584 ; заявл. 09.06.2004 ; опубл. 15.12.2005.
10. Bagherinia H. A theory of color barcodes / H. Bagherinia, R. Manduchi // Computer Vision Workshops (ICCV Workshops) : Міжнародна конференція, 6–13 листопада 2011 р. : тези доповідей. – Барселона : IEEE, 2011.
11. 2D Color Barcodes for Mobile Phones / [M. Querini, A. Grillo, et al.] // International Journal of Computer Science and Applications. – 2011. – Т. 8, № 1. – С. 135–155.
12. Bulan O. High capacity color barcodes using dot orientation and color separability / O. Bulan, V. Monga, G. Sharma // Media Forensics and Security : Симпозіум, 19–21 січня 2009 р. : тези доповідей. – Сан-Хосе : SPIE, 2009. – Т. 7254.
13. Parikh D. Localization and segmentation of a 2D high capacity color barcode / D. Parikh, G. Jancke // The IEEE Workshop on Applications of Computer Vision : Міжнародний семінар, 7–9 січня 2008 р. : тези доповідей. – Коппер-Маунтен : IEEE, 2008.
14. Патент США US8100330B2. Method for encoding and decoding data in a color barcode pattern / O. Bulan, V. Monga, G. Sharma ; заявник Xerox Corp. – № US12/436,456 ; заявл. 06.05.2009 ; опубл. 11.11.2010.
15. Hao T. COBRA: Color barcode streaming for smartphone systems / T. Hao, R. Zhou, G. Xing // Mobile systems, applications, and services (MobiSys'12) : XX Міжнародна конференція, 25–29 червня 2012 р. : тези доповідей. – Лой Вуд Бей : ACM, 2012. – С. 85–98.