

DOI: 10.20535/kpi-sn.2019.2.167521

UDC 004.627

M.V. Onai, O.K. Sulema*, A.I. Dychka
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine
*corresponding author: olga.sulema@pzks.fpm.kpi.ua

DATA ENCODING BASED ON TRICOLOR MATRIX BARCODES

Background. Nowadays, matrix barcodes in general and QR codes in particular are widely used in multiple applications, when it is important to input data into a computer system in a fast and secure way (both error-free and protected from third parties). Their popularity was significantly increased since the era of smartphones had begun, and so consequently, new use cases and new problems have arisen. Specifically, one of the important issues is representing more information as a barcode symbol.

Objective. The objective of the research is to develop a data barcoding method based on grayscale tricolor 2D barcode which enables representing more amounts of information with preservation of the area of a graphical symbol.

Methods. The research is based on both the coding theory and the multicolor barcode concept, which allow to increase amounts of data that can be encoded and represented by one matrix barcode symbol. The proposed BGW barcoding method uses a combination of data transformations between several notation systems with the purpose of data compression.

Results. The method of data encoding based on tricolor matrix barcodes is proposed and described. This method allows to compress input alphanumeric data and transform it into a 2D barcode symbol composed of BGW (black, gray, white) barcode patterns. Two use cases for the proposed method application are considered in the paper. The first use case concerns enhanced information input. The second one addresses the data protection issue.

Conclusions. The proposed BGW barcoding method enables compact representation of a textual data. One of the benefits of the proposed method lies in a possibility to not only encode data but to compress it. Due to data compression, more information can be represented keeping the same dimensions of the graphically coded symbol. Depending on the parameters, especially barcode symbol digital capacity, compression can be up to 20–25%.

Keywords: barcoding; data encoding; data compression; matrix barcode; tricolor barcode.

Introduction

Matrix, or 2D, barcodes have been effectively used to make data input into a computer system faster and more secure (both free error and protected from third parties) for decades, but popularity of 2D barcodes in general and QR codes in particular was significantly increased since the era of smartphones had begun. Multiple use cases in different fields of humans' activity have appeared, such as ID verification, shipping and receiving, mobile ticketing or document tracking [1], and these are only few examples of using barcodes, along with new use cases and, respectfully, new problems arising. Thus, one of the subjects of particular interest is encoding more information in the form of barcode. It can be easily achieved by extending the area of a graphical barcode symbol, however it could be an inappropriate solution for certain use cases, when size is substantial (e.g. barcodes on microcircuit).

Barcoding methods and barcodes themselves are a subject of study for many researchers. Thus, in [2], the authors propose a new approach to color barcode decoding that does not require a reference

color palette. According to their algorithm, groups of color bars are decoded at once, exploiting the fact that joint color changes can be represented by a low-dimensional space.

The authors of [3] present an approach for localization and segmentation of a 2D color barcode when it is read using computer vision techniques. They develop a progressive strategy to achieve high accuracy in diverse scenarios and computational efficiency.

In [4], a visible light communication (VLC) system for off-the-shelf smartphones entitled COBRA is presented. The system encodes information into specially designed 2D color barcodes. The authors developed a new barcode for COBRA, which is optimized for streaming between small-size screen and low-speed camera of smartphones.

The authors of the patent [5] propose to store information decoded from a bar-code as character-based data in an auxiliary field, e.g. a comment field, of an image file.

In [6], the authors proposed High Capacity Colored QR codes (HCC2D) as an alternative to a standard QR codes. The main idea authors

developed in the paper is to create a new 2D code which aims at increasing the space available for data, while preserving similar robustness, error correction and without losing compatibility with the original QR standard. Authors compared their new approach to Microsoft's High Capacity Color Barcode (HCCB) described in [7]. It is shown in the paper that HCC2D approach leads to a larger data density compared to QR at the price of a small computational overhead. Though the data density is slightly lower than in HCCB, HCC2D does not suffer from the problems in detection and alignment of the 2D code.

The author of [8] proposes a method of generating and decoding of two-dimensional color barcode, which includes a black and white configuration block that encodes configuration information about the barcode and a plurality of color data blocks that encodes data.

A method of high capacity color barcodes generation is proposed in [9]. This method operates by embedding independent data in two different printer colorant channels via halftone-dot orientation modulation.

In [10], the authors present a system and method for encoding and decoding data in a color barcode pattern using dot orientation and color separability. The authors assert that the method is robust against interseparation misregistration with a small symbol error rate.

The authors of [11] present a prototype for generating and reading the HCC2D code format on both PC (Linux and Windows platforms) and mobile phones (Android platform). The experimental results considering different operating scenarios and data densities in comparison with 2-dimensional barcodes are provided.

Thus, although there are a lot of solutions for barcoding, a problem of improving data representation as a barcode is still relevant and, among other issues, requires new approaches in data compressing.

Problem Statement

The subject of this study is a barcoding of data with compression.

A possible way to achieve increasing amount of information to be encoded with preservation of the area of a graphical symbol is to add more colors to a barcode and make it multicolor instead of black-and-white.

However, in case of using RGB palette there are few problems. In the first place, it will perceptibly augment overall complexity of calculations as

each of RGB matrices has to be processed. Secondly, multicolor barcodes can be effectively used in the digital form. Modern printers are highly advanced, however there still could be a problem of color rendering. Particular colors can be distorted when printed, with the result that barcode scanner will incorrectly process encoded data.

Grayscale tricolor 2D barcode, or Black-Gray-White barcode (BGW barcode), solves the problem of representing multicolor barcodes in the printed form. Because of using grayscale, a barcode consisted of black and white colors and shades of gray can be easily printed using grayscale printing mode. In this case, barcode scanner will analyze an intensity of gray color, which would enable to minimize the possibility of scanning error.

In this paper, we propose the data barcoding method that enables to encode more amounts of information with preservation of the area of a graphical symbol, based on grayscale tricolor 2D barcode.

Data BGW Barcoding Method

Let us define some fundamental notations. A *2D barcode symbol* is a set of barcode patterns that are densely spaced on a carrier in the form of matrix. A *barcode pattern* is graphical representation of s -digits ternary sequence of symbols.

A barcode pattern consists of s elements which are, in physical meaning, matrix cells. Each cell can be painted one of 3 colors; let us label them as white, gray and black. Consequently, maximum capacity of a 2D barcode symbol (i.e capacity of a set of all possible barcode patterns) will be $V_{\max} = 3^s$ barcode patterns.

Let us consider 2 types of barcode patterns: informational barcode patterns and auxiliary barcode patterns. Informational ones are used to present incoming alphanumeric sequence of symbols. Auxiliary ones are used to switch between modes, to indicate START and STOP symbols, scanner commands, etc.

Symbology Ω of a barcode is a set of all possible barcode patterns at fixed s . Let us consider the symbology Ω to be an *alphabet* Ω with capacity $P_{\Omega} = 3^s$. This alphabet consists of informational patterns Ω_{inf} and auxiliary patterns Ω_{aux} , i.e. $\Omega = \Omega_{inf} \cup \Omega_{aux}$. Capacity of informational patterns is $P_{\Omega_{inf}}$ and, respectively, capacity of auxiliary patterns is $P_{\Omega_{aux}}$. Thus, $P_{\Omega_{inf}} + P_{\Omega_{aux}} = 3^s$. Input information that will be transformed into

symbology Ω represents a sequence of symbols belonged to an alphabet A . The alphabet A consists of Roman alphabet, Cyrillic alphabet, digits, and special symbols and is a subset of alphabet ASCII, i.e. $A \subset ASCII$.

A. Theoretical Background

Let an entering alphanumeric sequence is as follows:

$$\mathbb{T} = t_1 t_2 \dots t_h \quad (1)$$

where t_i is an element of entering sequence, $t_i \in ASCII$, $i = 1, 2, \dots, h$ and h is sequence \mathbb{T} length.

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

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

$$\begin{cases} w_1 = t_1 t_2 \dots t_{i-1} \in L, \\ w_2 = t_i t_{i+1} \dots t_{h-i+1} \in D, \\ \dots \\ w_k = t_{h-i} t_{h-i-1} \dots t_h \in C. \end{cases}$$

Thus, the entering sequence acquires the following form:

$$\mathbb{T} = w_1 w_2 \dots w_k$$

where w_i is a subsequence of the entering sequence which contains elements of only one set of ASCII subsets. However, in the entering sequence \mathbb{T} the subsequences $w_1 w_2 \dots w_k$ can be situated in any order.

Let us consider a mathematical model of the entering sequence \mathbb{T} transformation to a compressed sequence \mathbb{U} .

The subsequence $w_i = t_1 t_2 \dots t_n$ is considered as n -digits vector in a notation with a base P_A . After compressing this subsequence transforms to m -digits vector in a notation with a base $P_{\Omega_{inf}}$ where $n > m$ and $P_A \leq P_{\Omega_{inf}}$. A sequence of j vectors $u_i = c_1 c_2 \dots c_m$ ensues so that the following compressed sequence \mathbb{U} is being obtained:

$$\mathbb{U} = u_1 u_2 \dots u_j. \quad (2)$$

The process of compressing a sequence of n adjacent symbols that belong to the alphabet A with a cardinality P_A reduces to transforming these symbols to m barcode patterns, which are symbols of the alphabet Ω_{inf} with cardinality $P_{\Omega_{inf}}$:

$$n(P_A) \rightarrow m(P_{\Omega_{inf}}) \quad (3)$$

where n is a number of symbols from A and m is a number of symbols from $P_{\Omega_{inf}}$.

Thus, the transformation (3) is a transformation of a number from one notation system to another one. It can be represented as following:

$$\sum_{i=0}^{n-1} \alpha_i P_A^i \rightarrow \sum_{j=0}^{m-1} \omega_j P_{\Omega_{inf}}^j$$

where α_i are corresponding codes of symbols from A and ω_j are corresponding codes of symbols from Ω_{inf} .

In this case, the entering sequence (1) will be compressed to the resulting sequence (2) if $n] \log_3 P_A [> m] \log_3 P_{\Omega_{inf}} [$ and $P_A^n - 1 \leq P_{\Omega_{inf}}^m - 1$ where $P_A^n - 1$ and $P_{\Omega_{inf}}^m - 1$ are quantitative equivalents of, correspondingly, maximal n -digits number in a notation P_A and maximal m -digits number in a notation $P_{\Omega_{inf}}$. These conditions can be represented as the following inequality system:

$$\begin{cases} P_A^n \leq P_{\Omega_{inf}}^m, \\ n] \log_3 P_A [> m] \log_3 P_{\Omega_{inf}} [\end{cases} \quad (4)$$

where $n] \log_3 P_A [$ is a length of an entering uncompressed sequence of symbols and $m] \log_3 P_{\Omega_{inf}} [$ is a length of a resulting compressed sequence.

Hence, the aim is to find such P_A for fixed $P_{\Omega_{inf}}$ that a maximal compressing will be guaranteed when a sequence of symbols is transforming to a 2D barcode symbol.

To assess compression degree, let us calculate a ratio of an entering sequence length to a compressed sequence length:

$$U_{P_{\Omega_{inf}}}^{(s)}(P_A) = \frac{n] \log_3 P_A [}{m] \log_3 P_{\Omega_{inf}} [}.$$

Hereinafter the ratio will be termed a *compression coefficient*.

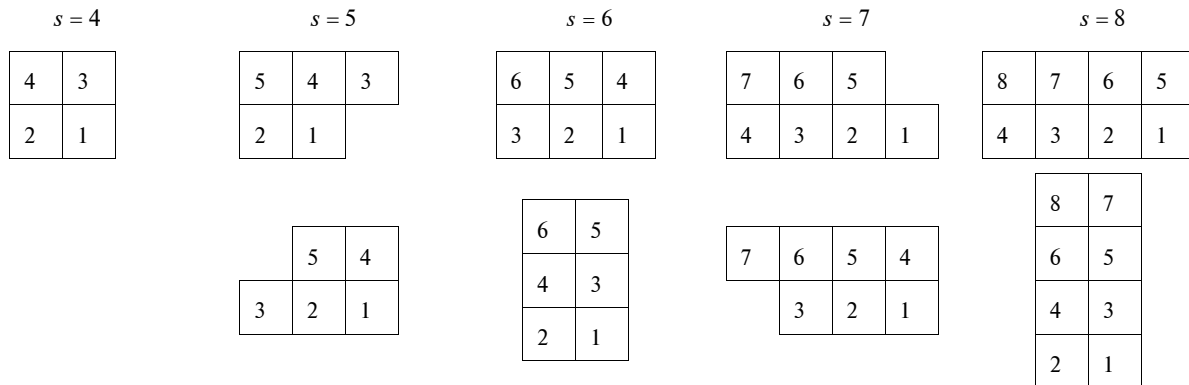


Fig. 1. Bits arrangement in a barcode pattern according to digit capacity s

B. BGW 2D Barcode Construction

As was stated at the beginning of the section, a barcode pattern is a graphical representation of s -digits ternary sequence of symbols, and a maximum capacity of a 2D barcode symbol is the capacity of a set of all possible barcode patterns which will be $V_{max} = 3^s$ barcode patterns.

Table 1 shows the correspondence between a parameter s value and maximal capacity of a 2D barcode symbol.

It is inadvisable to consider barcode patterns when $s < 4$ as such 2D barcode symbols have too small capacity to be applied to real-life problems. Thus, $s = 4, 8$ will be of practical value in further research.

Fig. 1 shows possible alternatives on how to graphically represent a barcode pattern in accordance with s value.

The procedure of transforming an entering alphanumeric sequence to a 2D barcode symbol consists of the following stages:

Table 1. Dependence of 2D barcode symbol maximal capacity on barcode pattern digital capacity

s value	3^s value	Maximal capacity of a 2D barcode symbol
4	3^4	81
5	3^5	243
6	3^6	729
7	3^7	2187
8	3^8	6561
9	3^9	19683
10	3^{10}	59049

I. Transformation of symbols from the entering sequence to a sequence of codes of corresponding symbols. At this stage, each symbol from the alphanumeric sequence is being replaced by a code corresponded to an order number of this symbol in the alphabet. As a result, we obtain a sequence of order numbers (Fig. 2).

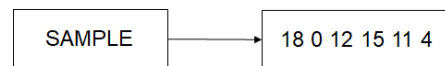


Fig. 2. First transformation of the entering sequence

II. Transformation of a decimal number obtained at the first stage from notation system P_A to notation system $P_{\Omega_{inf}}$. Applying the transformation (6), we obtain a number in notation $P_{\Omega_{inf}}$ (Fig. 3) at the second stage.

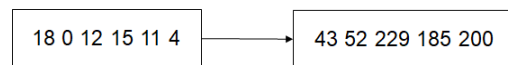


Fig. 3. Second transformation of the entering sequence

III. Transformation of the number in notation $P_{\Omega_{inf}}$ to a ternary number. At the third stage, we transform each number in notation $P_{\Omega_{inf}}$ to ternary system and obtain the ternary number (Fig. 4).

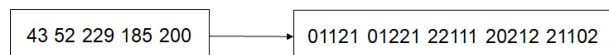


Fig. 4. Third transformation of the entering sequence

IV. Transformation of the ternary number to a barcode pattern. Finally, the number in ternary system can be represented in a form of matrix each cell of which is painted over one of three colors: white

for digit 0, gray for digit 1, and black for digit 2. At the end of this process, we obtain a barcode symbol representing the entering data in a graphical-coded form (Fig. 5).



Fig. 5. Final transformation of the entering sequence

Thus, transforming initial alphanumeric data to a barcode occurs in 5 consecutive steps.

Multiple barcode patterns transformed from initial alphanumeric sequences of characters form a 2D barcode symbol which, afterwards, can be located at a physical object.

C. Practical Implementation

Let us consider an example of the proposed method implementation.

A lot of public and private offices keep information about their customers, or patients in case of medical institution. Generally, this information contains such personal data as full name, date of birth, passport number, residential address. To prevent third parties from accessing to these data, they could be transformed to a 2D barcode symbol by following the steps stated in the previous subsection.

Let the entering alphanumeric sequence E is as follows: “OLEKSIY KLYMENKO, 01/01/1990, XX83629, 36 KHRESHCHATYK STR., 01044 KYIV, UKRAINE.” (all the data in this example are fictitious).

As it is quite a small amount of information, there is no need to use large s values which are more appropriate to encode, for instance, a medical history or any other body of data. Hence, let $s = 4$ in this example. Then $P_{\Omega_{inf}} = 70$ as $P_{\Omega} = 81$ and

$P_{\Omega_{aux}}$ has been chosen to be equal to 11 in this case.

As a pre-step for barcoding, we need to determine alphabets which will be used when encoding the entering sequence of symbols. To do this, inequality system (7) must be solved for integer values only. When $s = 4$, we obtain a set of solutions. The most useful integer solutions are presented in Table 2.

As we are going to encode Latin symbols, digits and some punctuation symbols (including space), the least possible cardinalities for each of these alphabets are 26, 10 and 4 correspondingly. However, if we analyze possible subsequences in these

particular textual data, it is evident that punctuation symbols can be included into Latin and digits alphabets to avoid a necessity to switch between alphabets just because of one not-letter-symbol. Hence, we need to look for alphabets with cardinality covering those punctuation symbols.

Table 2. Some integer solutions for $s = 4$

m	n	Cardinality P_A of alphabet A	Type of transformation $n \rightarrow m$	Compression coefficient $U_{70}^{(4)}(P_A)$
2	3	16	$3 \rightarrow 2$	1.125
3	5	12	$5 \rightarrow 3$	1.250
4	5	29	$5 \rightarrow 4$	1.250
	7	11	$7 \rightarrow 5$	1.313
5	6	34	$6 \rightarrow 5$	1.200
	9	10	$9 \rightarrow 5$	1.350
6	7	38	$7 \rightarrow 6$	1.167
	11	10	$11 \rightarrow 6$	1.375
7	8	41	$8 \rightarrow 7$	1.143
	9	27	$9 \rightarrow 7$	1.286
8	9	43	$9 \rightarrow 8$	1.125
	13	13	$13 \rightarrow 8$	1.219
9	10	45	$10 \rightarrow 9$	1.111
	11	32	$11 \rightarrow 9$	1.222
	16	10	$16 \rightarrow 9$	1.333
10	11	47	$11 \rightarrow 10$	1.100
	16	14	$16 \rightarrow 10$	1.200

While choosing, we must also observe compression coefficient and a type of transformation. Even though alphabet with cardinality $P_D = 14$ fits better to cover 4 punctuation symbols, it has worse both type of transformation $16 \rightarrow 10$ and compression coefficient $U_{70}^{(4)}(P_D) = 1,20$ than an alphabet with cardinality $P_D = 12$ has. So, in this case we can reduce punctuation symbols from 4 to 2 for alphabet of digits.

In general, these alphabets are being defined for each specific area of application (i.e., healthcare, public service, banking, etc.). In areas where large amounts of information must be stored and, consequently, encoded there could be a set of at least 3 alphabets, namely Latin letters, digits and punctuation symbols, including special characters. In some use cases ASCII both standard and extended can be also added.

In the final analysis, we choose the following alphabets to encode the data:

- alphabet L for Latin symbols, including space, full stop and comma, with cardinality $P_L = 29$, a type of transformation $5 \rightarrow 4$ and compression coefficient $U_{70}^{(4)}(P_L) = 1,25$;
- alphabet D of digits, including space and slash, with cardinality $P_D = 12$, a type of transformation $5 \rightarrow 3$ and compression coefficient $U_{70}^{(4)}(P_D) = 1,25$.

As soon as the pre-step is done, we need to follow 4 stages of data transforming to a 2D barcode symbol which have been stated in the previous subsection.

At the first stage, the entering sequence splits up into a set of subsequences of 5 symbols as there are transformations $5 \rightarrow 4$ and $5 \rightarrow 3$. When dividing, as soon as we reach the first digit, which does not belong to alphabet L , we insert a switch mark S_D and switch to alphabet D . Analogously, as soon as we reach a symbol that does not belong to alphabet D , we input a switch mark S_L and switch back to alphabet L .

As a result, the set of subsequences is as follows: { Oleks; iy_Kl;ymenk; o,_; S_D ;01/01; /1990; S_L ; ,_XX; S_D ; 83629 ; S_L ; , ; S_D ; _36_; S_L ; Khres; hc hat; yk_st; r.,_; S_D ; 01044; S_L ; _Kyiv; ,_Ukr; aine. }, where underscore means space.

Each symbol in the set is being replaced with a corresponding code that is an order number of this symbol in the alphabet.

At the second stage, each subsequence is being transformed from notation system with base 29 for alphabet L and base 12 for alphabet D to notation system with base 70:

```

14 11 4 10 18 → 29 46 21 15
8 24 26 10 11 → 18 18 50 51
24 12 4 13 10 → 50 24 50 63
14 28 26 → 2 40 12
29 → 29
0 1 10 0 1 → 0 45 19
10 1 9 9 0 → 42 67 2
12 → 12
28 26 23 23 → 23 67 58
29 → 29
8 3 6 2 9 → 35 6 49
12 → 12
28 → 0 28
29 → 29
11 36 11 → 3 68 63
12 → 12
10 7 17 4 18 → 21 11 15 14
7 2 7 0 19 → 14 41 39 21

```

```

24 10 26 18 19 → 50 18 40 41
17 27 28 26 → 1 19 29 28
29 → 29
0 1 0 4 4 → 0 25 30
12 → 12
26 10 24 8 21 → 54 26 60 33
28 26 20 10 17 → 59 44 35 59
08 13 4 27 → 0 42 5 38

```

At the third stage, each number in each obtained subsequence is being transformed to ternary number, so that the result is a sequence of ternary numbers: 1002 1201 0210 0120 0200 0200 1212 1220 1212 0220 1212 2100 1111 0110 1002 1200 0201 1120 2111 0002 0110 0002 0010 2111 2011 1002 1022 0020 1211 0110 0000 1001 1002 0010 2112 2100 0110 0210 0102 0120 0112 0112 1112 1110 0210 1212 0200 1111 1112 0001 0201 1002 1001 1002 0221 1010 0110 2000 0222 2020 1020 2012 1122 1022 2012 0000 1120 0012 1102.

Finally, the sequence obtained above is being represented in a form of matrix 7×10 where white color represents 0, gray 1 and black 2 (Fig. 6).

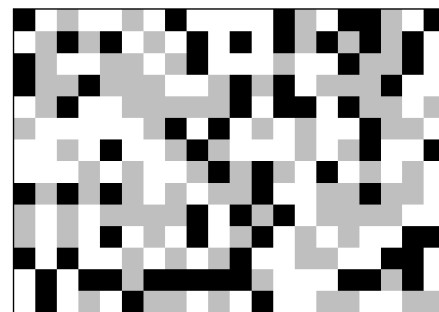


Fig. 6. The 2D barcode symbol containing the entering alphanumeric sequence E

Thus, the entering alphanumeric sequence containing private data has been transformed into the 2D barcode symbol. Its compressing degree is equal to 17%. Taking into account relatively small amount of the initial data, we consider that the obtained compression is satisfactory.

Grayscale Tricolor Barcoding Use Cases

A. Enhanced Information Input

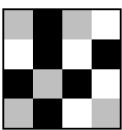

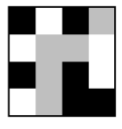
Initially barcodes have been used for quick and errorless input of information into a computer system. Today barcodes in general and QR-codes specifically are widely spread in advertising as an alternative URL to make a website access faster and more accurate, especially when a web-address is long and complex.

Quick and error-free data input is particularly vital in automated fields of activity, such as manufacturing. Even though automation have embraced most of industrial processes, people still take an active part in controlling them by entering specific commands into a computer system.

BGW barcode enables holding larger amounts of information in a graphical symbol, so that more computer instructions can be represented by a single symbol.

Similarly to the data protection approach, sets of commands for industrial equipment are presented in the form of BGW barcode. The printed barcodes will form a so-called command sheet (Table 3). Such a command sheet might contain independent commands, command sets or executable code, depending on a field of application. For instance, command sheet approach can be used for installing drivers and software or customizing equipment for specific purposes.

Table 3. Example of a command sheet

Barcode of a command set	Set of commands
	Command set 1: Command 1.1 Command 1.2 ... Command 1.m1
	Command set 2: Command 2.1 Command 2.2 ... Command 2.m2
...	...
	Command set N: Command 3.1 Command 3.2 ... Command 3.mN

A command sheet can be located at the equipment, so that technical personnel can easily read it with a scanner and enter these commands into a computer system instead of using keyboard and control keys.

The proposed information technology for enhanced information input in the industrial area is presented at Fig. 7.

Such approach ensures quick and error-free input of complex data or set of instructions preventing a possibility of incorrect functioning of the technical system.

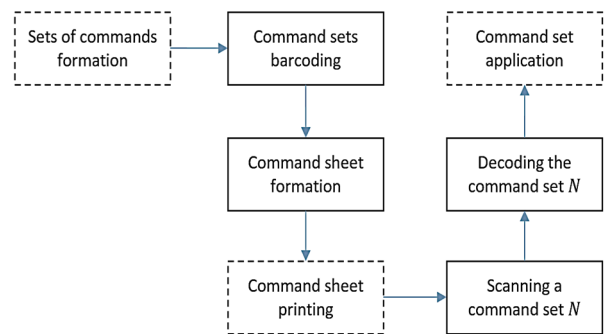


Fig. 7. Information technology for industrial application

B. Textual Data Protection

Nowadays, paperless information technologies are widely used. In many use cases information exists only in electronic form. Nevertheless, printed textual documents are still in active use and it is very likely they will keep their role for a long time.

Information presented as a printed textual document frequently has confidential character. Names, addresses of residence, birthdates, marital status – all these data should be kept in a secured way. However, to protect a printed document from unauthorized access is much harder than to ensure information security in a computer system where a document security management system can be used and both cryptography and steganography can be applied [12, 13].

Usually, the protection of printed documents is ensured by certain organizational procedures protecting information from access of unauthorized persons by limitation of physical presence of such people in offices as well as labeling documents with ‘for official use only’ mark. However, this approach cannot guarantee data protection.

The proposed approach for textual data protection is based on the fact that the human brain is not able both to recognize and to memorize data presented as a barcode. It is supposed that a textual document protected according to this approach can be visualized and perceived by a user only on a PC screen. When the document is printed, it is presented as a 2D barcode. Thus, the information technology we propose consists in the following (Fig. 8).

A computer system to be used for a textual documentation preparation includes a PC, a scanner, and a printer along with a special software.

A user uses PC for creation of a document which can contain a text in a certain language, e.g. in English. The document might be saved at the local data storage and it can be opened any time on

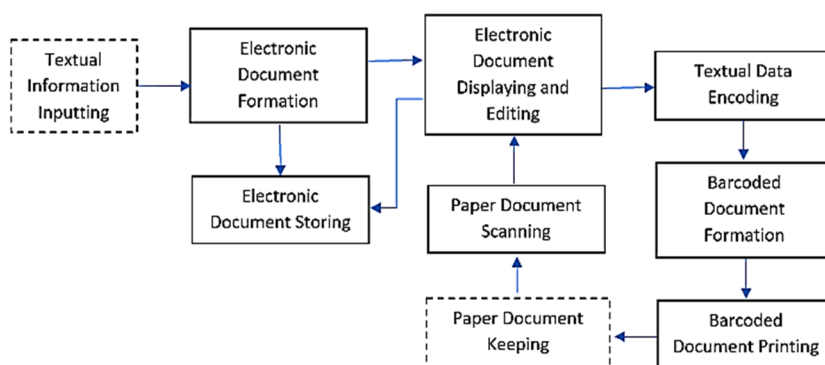


Fig. 8. Information technology for textual data protection

this PC. When the document is opened, it is visualized in a usual way. At the same time, when the user prints the document, it is being converted into a 2D barcode and only after this transformation it might be printed on a paper. The printing of non-barcode data is forbidden by the special software pre-installed at PC.

To see the printed textual document in a readable form, the user must scan the document. As the result of scanning, the text is displayed in the original form at the PC screen.

Conclusions

The proposed BGW barcoding method enables compact representation of a textual data. One of the benefits of the proposed method lies in a possibility to not only encode data but to compress it. Due to

data compression, we can encode more information keeping the same dimensions of the graphical-coded symbol.

In the example demonstrated above, the compression degree comes to 17%. Depending on the parameters, especially barcode symbol digital capacity, compression can be increased up to 20–25%.

The proposed method has a wide application. In particular, it can be used for enhanced information input. Another promising application area is textual data protection.

In this regards, we propose the information technology, which allows to prevent unauthorized access to printed documents of confidential nature. The data barcoding method we propose in the paper allows to protect private data by transforming it from its initial textual representation into a graphical form of a 2D barcode symbol.

The BGW barcoding method proposed in this paper has its potential to further research. As we can define any alphabet to be used when encoding alphanumeric sequence of symbols, data in any language could be protected by transforming into a 2D barcode symbol. Thus, software developed on the basis of the data barcoding method can be easily extended for any language, including not-Latin-based ones, such as Korean, Japanese, Chinese, Georgian, etc.

References

- [1] *Barcode Scanning Use Cases* [Online]. Available: <https://www.scandit.com/use-cases/>
- [2] H. Bagherinia and R. Manduchi, "A theory of color barcodes", in *IEEE Int. Conf. Comp. Vision Workshops*, 2011, Nov. 6–13, pp. 806–813. doi: 10.1109/ICCVW.2011.6130335
- [3] D. Parikh and G. Jancke, "Localization and segmentation of a 2D high capacity color barcode", in *IEEE Workshop on Applications of Computer Vision*, 2008, Jan. 7–9, pp. 1–6. doi: 10.1109/WACV.2008.4544033
- [4] T. Hao *et al.*, "COBRA: Color barcode streaming for smartphone systems", in *Proc. 10th Int. Conf. Mobile Systems, Applications, and Services*, 2012, Jun. 25–29, pp. 85–98. doi: 10.1145/2307636.2307645
- [5] J. Barrus and G.J. Wolff, "Embedding barcode data in an auxiliary field of an image file", U.S. Patent 7 150 399, Dec. 19, 2006.
- [6] A. Grillo *et al.*, "High capacity colored two dimensional codes", in *Proc. Int. Multiconf. Computer Science and Information Technology*, 2010, Oct. 18–20, pp. 709–716.
- [7] *High Capacity Color Barcodes (HCCB)* [Online]. Available: <http://research.microsoft.com/en-us/projects/hccb/>
- [8] P. Catrone, "Two-dimensional color barcode and method of generating and decoding the same", U.S. Patent 7 478 746, Jan. 20, 2009.
- [9] O. Bulan *et al.*, "High capacity color barcodes using dot orientation and color separability", *Media Forensics and Security*, vol. 7254, 2009. doi: 10.1117/12.807742
- [10] O. Bulan *et al.*, "Method for encoding and decoding data in a color barcode pattern", U.S. Patent 8 100 330, Jan. 24, 2012.
- [11] M. Querini *et al.*, "2D color barcodes for mobile phones", *Int. J. Comp. Sci. Appl.*, vol. 8, no. 1, pp. 136–155, 2011.

- [12] S. Chourasia and K.N. Singh, "An efficient hybrid encryption technique based on DES and RSA for textual data", in *Information Systems Design and Intelligent Applications*, S. Satapathy et al., eds. New Delhi: Springer, 2016. pp. 73–80. doi: 10.1007/978-81-322-2755-7_9
- [13] M.J. Khami, "Transmitting security enforcement by text encrypting and image hiding technique using combined encrypt/hide keys", *Int. J. Engineering and Manufacturing*, vol. 1, pp. 1–15, 2018. doi: 10.5815/ijem.2018.01.01.

М.В. Онай, О.К. Сулема, А.І. Дичка

КОДУВАННЯ ДАНИХ НА ОСНОВІ ЗАСТОСУВАННЯ ТРИКОЛІРНИХ МАТРИЧНИХ ШТРИХКОДІВ

Проблематика. На сьогодні матричні штрихові коди в цілому і QR-коди зокрема широко використовуються у багатьох застосунках, коли є важливим вводити дані в комп'ютерну систему швидко і безпечно (з точки зору як відсутності помилок при введенні, так і обмеженості доступу сторонніх осіб). З початком епохи смартфонів популярність матричних штрихових кодів значно зросла, в результаті чого виникли нові сценарії використання та нові проблеми. Зокрема, однією з важливих проблем є подання у вигляді штрихкоду більших обсягів інформації.

Мета дослідження. Метою дослідження є розроблення методу кодування даних, що дає можливість перетворювати на матричний штриховий код у градаціях сірого більші обсяги даних зі збереженням при цьому розміру графічного символу.

Методика реалізації. Проведене дослідження ґрунтується на теорії кодування та концепції багатоколірного штрихового кодування, що дає змогу збільшувати кількість інформації, яка може бути представлена одним штрихкодним знаком. Запропонований метод кодування BGW (скорочення від англ. "чорний-сірий-білий") використовує послідовність перетворень даних між різними системами числення з метою ущільнення даних.

Результати дослідження. Запропоновано та описано метод кодування даних, що ґрунтується на триколірних матричних штрихових кодах. Цей метод дає можливість ущільнювати вхідні алфавітно-цифрові дані та подавати їх у вигляді матричної штрихкової позначки, штрихкодіві знаки якої складаються з чорних, сірих і білих комірок. Розглянуто два сценарії використання запропонованого методу. Перший сценарій передбачає використання вдосконаленого способу введення інформації. Другий сценарій стосується захисту даних.

Висновки. Запропонований метод кодування BGW забезпечує компактне подання текстових даних на фізичному носії. Однією з переваг запропонованого методу є можливість не лише кодувати дані, але й ущільнювати їх. Унаслідок ущільнення даних можна подавати на носії більше інформації, не змінюючи при цьому розміри графічно-кодованого символу. Залежно від вибраних параметрів, зокрема ємності штрихкової позначки, ущільнення даних може становити до 20–25 %.

Ключові слова: штрихкодування; кодування даних; ущільнення даних; матричні штрихові коди; триколірні штрихові коди.

Н.В. Онай, О.К. Сулема, А.І. Дичка

КОДИРОВАНИЕ ДАННЫХ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ТРЕХЦВЕТНЫХ МАТРИЧНЫХ ШТРИХКОДОВ

Проблематика. На сегодняшний день матричные штриховые коды в целом и QR-коды в частности широко используются во многих приложениях, когда важно вводить данные в компьютерную систему быстро и безопасно (с точки зрения как отсутствия ошибок при введении, так и ограниченности доступа посторонних лиц). С началом эпохи смартфонов их популярность значительно повысилась, в результате чего появились новые сценарии использования и новые проблемы. В частности, одной из важных проблем является кодирование в виде штрихового кода больших объемов информации.

Цель исследования. Целью исследования является разработка метода кодирования данных, который позволял бы преобразовывать в трехцветный штриховой код в градациях серого большие объемы данных с сохранением при этом изначальных размеров графического символа.

Методика реализации. Проведенное исследование основывается на теории кодирования и на концепции многоцветных штриховых кодов, что позволяет увеличивать количество информации, которая может быть представлена одним штрихкодным знаком. Предложенный метод BGW (сокращение от англ. "черный-серый-белый") использует последовательность преобразований данных из одной системы счисления в другую с целью сжатия данных.

Результаты исследования. Предложен и описан метод кодирования данных, который основывается на трехцветных матричных штриховых кодах. Этот метод позволяет сжимать исходные алфавитно-цифровые данные и преобразовывать их в матричный штрихкодный символ, штрихкодные знаки которого состоят из черных, серых и белых ячеек. Рассмотрены два сценария использования предложенного метода. Первый сценарий предусматривает использование усовершенствованного способа введения информации. Второй сценарий связан с задачей защиты данных.

Выводы. Предложенный метод кодирования BGW обеспечивает компактное представление текстовых данных на физическом носителе. Одним из преимуществ предложенного метода является возможность не только кодировать данные, но и сжимать их. В результате сжатия данных можно представлять на носителе больше информации, не изменяя при этом размеры графически-кодированного символа. В зависимости от выбранных параметров, в частности емкости штрихкодного символа, сжатие данных может достигать 20–25 %.

Ключевые слова: штрихкодирование; кодирование данных; сжатие данных; матричные штриховые коды; трехцветные штриховые коды.

Рекомендована Радою
факультету прикладної математики
КПІ ім. Ігоря Сікорського

Надійшла до редакції
26 березня 2019 року

Прийнята до публікації
25 квітня 2019 року