DOI 10.34658/jack.2019.27.2.17-38

JOURNAL OF APPLIED
COMPUTER SCIENCE
Vol. 27 No. 2 (2019), pp. 17-38

Black-Gray-White Barcode Based on Error Correction Data Encoding

Ivan Dychka, Olga Sulema

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Faculty of Applied Mathematics
Prosp. Peremohy 37, 03056 Kyiv, Ukraine
dychka@pzks.fpm.kpi.ua, olga.sulema@pzks.fpm.kpi.ua

Abstract. *The paper presents an approach based on using barcodes with three color gradations (white, gray, and black) instead of traditional black and white barcodes. Such three-color barcodes can be produced and read by using the same equipment as usual black and white barcodes, but due to the third color they have a higher information density.*

It is shown in the paper that in order to ensure high reliability and accuracy of barcoded information reading, it is necessary to encode the data with a Reed-Solomon correction code which is capable of correcting multiple errors.

The method of data encoding and decoding in the Galois field $GF(3^m)$, where m is the degree of an irreducible polynomial, is proposed. The noise immunity of the black-gray-white barcode capable of multiple errors correction based on Reed-Solomon code is analyzed and discussed as well.

Keywords: *barcoding, multicolored barcodes, Galois fields, noise immunity, error correction.*

1. Introduction

Modern management systems used in production, transportation, trade, service, and other areas require accurate and relevant information about the status of accounting and tracking objects. This requirement has led to the rapid development of automatic identification technologies, which allow to input data into a computer system by automatic reading from the object. Automatic identification ensures that the computer system receives real-time data and provides fast and accurate information input. Barcode technology has become especially widespread among various types of automatic identification.

Barcoding is a method of representing information by using simple graphical shapes: circle, square, rectangle, bar, polygon (hexagon, octagon), etc. [1]. Barcoding provides an optical way of data reading, including remote reading. In a barcode, information is provided by combinations of colored elements. The main advantages of data barcoding are its relative simplicity and low cost of producing barcode labels.

The analysis of existing barcodes shows that most of the barcoding methods allow to represent only small amounts of information (up to 3Kb) and, thus, a barcode is mainly used as a key to access more information stored in the computer system. Besides, most of existing barcodes have a low information density due to using only two colors (usually, black and white), which does not allow to store desirable amounts of information on an object surface.

In modern systems of automatic identification, the following demands to the barcode are made: 1) miniaturization of a barcode (a limited area is allocated for a barcode image on an object surface); 2) significant increase in the barcode capacity without changing its geometric dimensions; 3) providing a barcode with database features (the barcode should contain complete information about the object and not just be a key to access information); 4) barcode processing should be performed with high reliability.

Thus, there is an important scientific and technical problem of developing new methods for reliable storing large amounts of information presented as a barcode that can be located on a limited area on the carrier, as well as ensuring high reliability of the barcoded data reading.

In this paper we consider black-gray-white matrix barcodes (BGW-Code) as an alternative to classic black and white barcodes. The approach proposed in the paper allows to increase information density of barcodes without changing geometric sizes of a barcode label. It requires low-cost ordinary equipment for the production of barcodes, namely, a printer and a scanner. To achieve the high accuracy of data reading, we propose the three-color barcoding method based on error-correction codes.

The rest of this paper is organized as follows. In Section 2, the related work is presented. The approach to the control of distortions in barcodes is given in Section 3. The encoding method based on error correction codes is presented in Section 4. In Section 5, the algorithm of error-correction data decoding is defined. The experimental part of the research is highlighted in Section 6. The obtained results are discussed in Section 7. Finally, Section 8 concludes the paper and presents future research work.

2. Related work

In recent years, several interesting researches on multicolor barcoding as well as on error correction in barcodes have been presented.

[2] presents CodeCube, a multi-layer color barcode developed for mobile society applications. The authors adopt eight colors to implement a three-layer information capacity barcode through exploiting the digital color image red (R), green (G), and blue (B) channels. Due to this approach, CodeCube is used to transmit three independent data together. It adapts the barcode version size to deal with the problem caused by different lengths of the three independent data.

The authors of [3] describe a method of high capacity color barcodes generation, which operates by embedding independent data in two different printer colorant channels through halftone-dot orientation modulation.

In [4] a new approach to color barcode decoding is presented. The main idea is that it does not require a reference color palette. According to the algorithm described in the paper, groups of color bars are decoded at once, taking proper account of the fact that joint color changes can be represented by a low-dimensional space.

The authors of the patent [5] propose the way of storing data decoded from a barcode as character-based data in an auxiliary field of an image file.

High Capacity Colored QR codes are presented as an alternative to standard QR codes in [6]. It is proposed to create a new 2D code aimed at increasing the space available for data along with preserving similar robustness and error correction. HCC2D approach tends to a larger data density at the price of a small computational overhead, which makes it more preferable in comparison with Microsoft's

High Capacity Color Barcode (HCCB) [7] and standard QR codes.

In patent [8] the authors suggest both a system and the method for encoding and decoding data in a color barcode pattern using dot orientation and color separability. They claim that the proposed method is proof against interseparation misregistration with a small symbol error rate.

[9] presents JAB Code, a high-capacity 2D color bar code, which can encode more data than traditional black/white (QR) codes. It is a color two-dimensional matrix symbology whose basic symbols are made of colorful square modules arranged in either square or rectangle grids. The authors claim that JAB Code is able to encode from small to large amounts of data.

In [10], a prototype for generating and reading the HCC2D code format on both PC and mobile phones is presented. The authors provide the experimental results obtained with different operating scenarios and data densities and compare them with 2-dimensional barcodes.

A method of generating and decoding a two-dimensional color barcode is proposed by the authors of patent [11]. It includes a plurality of color data blocks, which encodes data, and a black and white configuration block, which encodes configuration information about the barcode.

The authors of [12] present an approach for localization and segmentation of a 2D color barcode when it is read using computer vision techniques. A progressive strategy of achieving high accuracy in diverse scenarios and computational efficiency.

In [13] the authors propose a security mechanism improvement method by applying the error correction with random segmentation (ECRS) technique. The proposed method has two advantages when compared with the traditional Reed-Solomon error correction coding algorithm. Due to the paper, the ECRS performs random segmentation using multiple redundant codes to protect data. At the same time, it protects the code from being falsified by using encryption keys.

A new 2D barcode encoding and decoding method is presented in [14]. In the proposed method, the authors divide the 2D code data area into several blocks and do not use for encoding the blocks where many of the bits in the block have high probabilities of bit errors. The decoder determines the skipped blocks from the image itself.

The analysis of these and other similar researches allows us to conclude that the development of new multicolor barcodes has a potential for further increasing of data representation density, however, these barcodes need to be armed with the error-correcting feature. There are several approaches aimed at providing noise immunity for barcodes, however, new multicolor barcodes are not mostly supported with reliable mechanisms for proper noiseless encoding. Thus, it is an important subject to study.

3. Control of distortions in barcodes

3.1. Structural organization of a barcode symbol

Barcode symbol (BC symbol, barcode) contains barcode patterns. Barcode pattern (BC pattern) is a minimal structural component of a barcode symbol, which consists of s elements (cells). BC patterns are located on a carrier in the form of a rectangle matrix.

An example of a barcode is shown in Figure 1. Figure 2 shows elements arrangement in a BC pattern according to digit capacity s.

Figure 1: An example of BGW-Code symbol

The use of three-color gradations allows the maximum capacity V_{max} of the barcode to be up to 3^s barcode patterns. Table 1 shows the dependence between the maximum capacity of a barcode (a barcode symbol, BC symbol) and the digit capacity s of a barcode pattern.

The barcode pattern consisted of s elements corresponds to s-digit ternary vector.

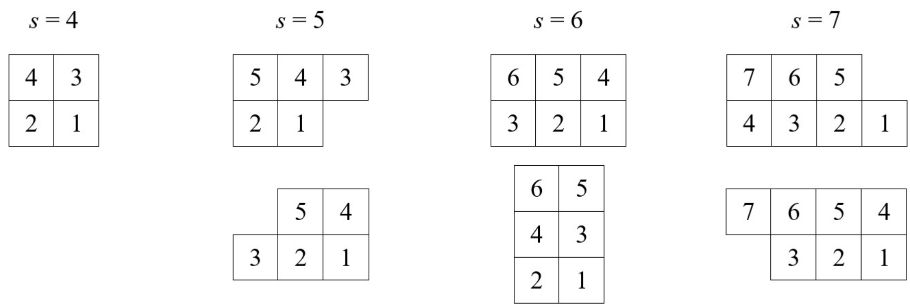

Figure 2: Bits arrangement in a barcode pattern according to digit capacity s

Table 1: Dependence between the maximum capacity of a barcode symbol and the digit capacity s of a barcode pattern

s	3^s	**Maximum capacity of a barcode symbol, V_{max} barcode patterns**
4	3^4	81
5	3^5	243
6	3^6	729
7	3^7	2187
8	3^8	6561
9	3^9	19683

3.2. Choosing an error correction code to control the data

An essential part of any barcode exploitation is ensuring its noise immunity and therefore, increasing the reliability of information stored as a barcode. This feature is highly important in any field, where barcodes are being used, in order to prevent data loss, both partly and fully, and consequently, to ensure that barcode performs its functions.

To assure barcode noise immunity, error-correcting codes [15, 16] can be used. Such codes are capable to correct errors in encoded data.

There is a number of reasons why barcode distortion occurs. First of all, it can happen at the stage of producing a barcode: (1) defects of barcode carrier, surface imperfections, low quality of barcode material, wrong choice of material, etc.; (2) low quality of used inks, peculiarity of ink cohesion with a carrier such as spreading or coagulation of ink, non-contrast image, etc.; (3) low quality and defects of printing due to failure of a printing device. Secondly, during storing and transporting a barcode symbol, its mechanical damage or deformations can occur causing errors. Other reasons are the dirtying of either a whole barcode symbol or its part and discoloration due to carrier and ink aging.

To ensure reliable reproduction of barcode data, it is necessary to encode the alphanumeric sequence with error-correcting codes which allow to correct distortions of two types: errors and erasures. An error means converting one barcode pattern into another one due to barcode graphical elements damage. An erasure means the case when a decoder is unable to correlate a barcode pattern to be decoded with any existing pattern because it does not exist in a barcode symbolism.

Among error-correcting codes [15] there is a number of codes able to fix multiple errors, for example, Bose-Chaudhuri-Hocquenghem code, Reed-Solomon code, Justesen code, Goppa code, etc. In this research, we use the Reed-Solomon code. The choice of this code can be justified due to several important factors.

First of all, most error-correcting codes are binary, while Reed-Solomon code is a multiple-valued code that operates with not bits but symbols belonged to a multiple-valued alphabet.

In BGW-Code such multiple-valued symbols are vectors that belong to an alphabet $\{0, 1, 2, ..., 3^s - 1\}$ and, in turn, they correspond to the barcode patterns. It allows us to correct the whole vector (i.e. barcode pattern) instead of correcting its separate parts (i.e. tricolor elements).

It is also important that Reed-Solomon codes have minimal redundancy. Finally, a procedure of encoding/decoding when using the Reed-Solomon code has smaller computational complexity in comparison with other error-correcting codes. These features make Reed-Solomon codes the most relevant to the stated problem.

Reed-Solomon code ensures fixing distortions of two types: errors and erasures. In order to strengthen the error-correcting effect, it is desirable to disperse adjacent barcode patterns throughout a barcode symbol while producing a barcode image.

If minimum code distance is equal to d^*, then any configuration of δ erasures can be recovered if $d^* \geq \delta + 1$. On the other hand, an error-correcting code with minimum code distance d^* allows to correct any configuration of γ errors and δ erasures assuming that

$$d^* \geq 2\gamma + 1 + \delta. \tag{1}$$

Let us consider the Reed-Solomon code with minimum code distance $d^* = n-k+1$, where n is an overall length of a codeword and k is a number of information symbols in a codeword.

Denoting $n - k$ by r, where r is a number of check symbols in codewords, we obtain

$$r \geq 2\gamma + \delta. \tag{2}$$

Consequently, if we need to correct, for instance, $\gamma = 3$ errors and $\delta = 4$ erasures, codewords of the Reed-Solomon code must contain at least $r = 10$ vectors of distortion correction (check symbols), as $2\gamma + \delta = 2 \cdot 3 + 4 = 10$.

Let one of $\Lambda + 1$ levels of distortion correction $(0, 1, 2, ..., \Lambda)$ be used depending on the quality of a barcode image carrier, quality of printing, etc. If it is a zero-level distortion correction, a BC symbol does not contain distortion correction vectors and all BC patterns are information ones.

Let us assume that level Λ of distortion correction guarantees fixing any distortion configuration which satisfies inequation

$$2\gamma + \delta \leq 2^{\Lambda} - 1. \tag{3}$$

However, other options for defining a ratio between Λ and quantities γ and δ are also possible. For example, $2\gamma + \delta \leq 2^{\Lambda+1} - 2$, $2\gamma + \delta \leq 2^{\Lambda+1}$, $\gamma + \delta \leq 2^{\Lambda+1}$ etc.

The correction level is being chosen depending on the capacity of a BC symbol (a greater number of BC patterns must correspond to a higher level of correction) and the environment where a barcode image scanning occurs.

Higher levels of distortion correction are useful if it is likely that a barcode image is considerably damaged. Lower levels of distortion correction are useful if we know in advance that BC symbol damage is unlikely.

In practice, the maximum level of distortion correction can be defined from the ratio $2\Lambda - 1 = \frac{2}{3}k$. The number of control vectors in codewords, in this case, totals two thirds of a number of information vectors k. However, if it is necessary, the number of control vectors can be increased.

4. Noiseless data encoding

Let us consider a BC pattern as a minimum information unit. In this case, the symbolism Ω of the BGW-Code, which is a set of all possible BC patterns, consists of $P_\Omega = 3^s$ BC patterns. The informational and auxiliary BC patterns, i.e. vectors, belong to Galois field $GF(P_\Omega)$.

The generator polynomial of the Reed-Solomon code is as follows:

$$g(x) = (x - \alpha)(x - \alpha^2)...(x - \alpha^r) \quad (4)$$

where α is a primitive element of the field $GF(P_\Omega)$ and r is a number of vectors of distortion correction in codewords.

The primitive element α is a number, the powers of which (from 0 to $n - 1$) with respect to the base of irreducible polynomial yield all nonzero elements of the field. If $P_\Omega = 729$, then $\alpha = 3$.

Let us extent $GF(3^6)$ over the field $GF(3)$. To obtain elements of the field $GF(3^6)$, we must use an irreducible polynomial of the power $s = 6$, which coefficients are elements of the field $GF(3)$ [17]. For instance, $m(x) = x^6 + x + 2$. Altogether, there are 60 irreducible polynomials of the power six and any of them can be chosen to obtain the field $GF(3^6)$.

The elements of the extension $GF(3^6)$ over the field $GF(3)$ can be presented in a form of: (1) nonnegative and negative power α, (2) polynomial of α, (3) ternary number, and (4) decimal number (see Table 2).

Table 2: Different forms of presentation of elements of the field $GF(3^6)$ modulo the irreducible polynomial $m(x) = x^6 + x + 2$

Power α	**Negative power α**	**Polynomial of α**	**Ternary number**	**Decimal number**
0	0	0	000000	0
α^0	α^{-728}	1	000001	1
α^1	α^{-727}	α	000010	3
α^2	α^{-726}	α^2	000100	9
α^3	α^{-725}	α^3	001000	27
...	...	...	...	...
α^5	α^{-723}	α^5	100000	243
...	...	...	...	...
α^7	α^{-721}	$2\alpha^2 + \alpha$	000210	21
α^8	α^{-720}	$2\alpha^3 + \alpha^2$	002100	63
...	...	...	...	...
α^{11}	α^{-717}	$\alpha^5 + \alpha + 2$	100012	248
α^{12}	α^{-716}	$\alpha^2 + \alpha + 1$	000111	13

Continued on the next page

Table 2: *Continued from the previous page*

Power α	**Negative power** α	**Polynomial of** α	**Ternary number**	**Decimal number**
...	...	...	...	...
α^{15}	α^{-713}	$\alpha^5 + \alpha^4 + \alpha 3$	111000	351
...	...	...	...	...
α^{22}	α^{-706}	$\alpha^4 + \alpha^2 + 2\alpha$	010120	96
...	...	...	...	...
α^{34}	α^{-698}	$2\alpha^5 + 2\alpha^4 + 2\alpha^3 + \alpha^2 + \alpha + 1$	222111	715
...	...	...	...	...
α^{44}	α^{-684}	$\alpha^5 + \alpha^4 + 2\alpha^2 + \alpha + 2$	110212	347
...	...	...	...	...
α^{370}	α^{-358}	$\alpha + 2$	000012	5
α^{371}	α^{-357}	$\alpha^2 + 2\alpha$	000120	15
...	...	...	...	...
α^{377}	α^{-351}	$2\alpha^3 + 2\alpha^2 + 2\alpha$	002220	78
...	...	...	...	...
α^{585}	α^{-143}	$2\alpha^5 + 2\alpha^4 + 2\alpha^2 + 2\alpha$	220220	672
...	...	...	...	...
α^{726}	α^{-2}	$\alpha^5 + \alpha^4 + 1$	110001	325
α^{727}	α^{-1}	$\alpha^5 + 1$	100001	244

Various forms of the element presentation can be used for different purposes. The polynomial form is useful for adding the field elements, for example, $243+347 = \alpha^5+\alpha^{44} = \alpha^5+(\alpha^5+\alpha^4+2\alpha^2+\alpha+2) = 2\alpha^5+\alpha^4+2\alpha^2+\alpha+2 = 210212_3 = \alpha^{589} = 590$ (coefficients are adding up by digits modulo 3). The nonnegative power form can be efficiently used for the operation of multiplication, for instance, $13 \cdot 96 = \alpha^{12} \cdot \alpha^{22} = \alpha^{34} = 409$. To fulfil the operation of division, it is reasonable to present both a dividend and a divisor in the form of a nonnegative power, after which the divisor should be transformed to the negative power form, for example, $248 \div 27 = \alpha^{11} \div \alpha^3 = \alpha^{11} \cdot \alpha^{-3} = \alpha^8 = 63$.

If the correction level $\Lambda = 2$ is chosen, then according to (2) and (3) we obtain $r = 3$. According to (4), a generator polynomial $g(x)$ is obtained as follows:

$$g(x) = (x-\alpha)(x-\alpha^2)(x-\alpha^3) = (x+2\alpha)(x+2\alpha^2)(x+2\alpha^3) = x^3 + (2\alpha^3 + 2\alpha^2 + 2\alpha)x^2 + (\alpha^5+\alpha^4+\alpha^3)x+\alpha+2 = x^3+\alpha^{377}x^2+\alpha^{15}x+\alpha^{370} = x^3+78x^2+351x+5.$$

Note that the identities $-1 \equiv 2$ and $-2 \equiv 1$ are true in the field $GF(3)$.

Consequently, when $r = 3$, the generator polynomial $g(x)$ in the field $GF(3^6)$ is equal to:

$$g(x) = x^3 + 78x^2 + 351x + 5. \tag{5}$$

In the general case, after opening the brackets (4) the generator polynomial is of the form $g(x) = \sum_{j=0}^{r} g_j x^j$, $g_r = 1$.

The informational vector $A = (a_k, a_{k-1}, ..., a_1, a_0)$, $a_i \in \{0, 1, 2, ..., 3^s - 1\}$ can be considered as an equivalent to the informational polynomial:

$$a(x) = \sum_{i=0}^{k-1} a_i x^i. \tag{6}$$

For example, an input sequence consisted of six vectors 403 470 262 436 720 302 corresponds to a polynomial $a(x) = 403x^5+470x^4+262x^3+436x^2+720x+302$.

The data encoding with Reed-Solomon code consists in multiplying the informational polynomial $a(x)$ by the generator polynomial $g(x)$ and obtaining the codeword $C = (c_{n-1}, c_{n-2}, ..., c_1, c_0)$, which corresponds to a code polynomial:

$$c(x) = \sum_{i=0}^{n-1} c_i x^i. \tag{7}$$

The coefficients c_i of the codeword C correspond to BC patters to be placed on a carrier.

If $r = 3$ that corresponds to the generator polynomial (5), then a code polynomial is as follows:

$$c(x) = a(x)g(x) = (403x^5 + 470x^4 + 262x^3 + 436x^2 + 720x + 302)(x^3 + 78x^2 + 351x + 5) = 403x^8 + 697x^7 + 179x^6 + 396x^5 + 597x^4 + 445x^3 + 526x^2 + 506x + 686.$$

It means that the input sequence of codewords 403 470 262 436 720 302 corresponds with a code sequence 403 697 179 396 597 445 526 506 686, where either concurrent correction of one error and one erasure, or correction of three erasures is provided.

Thus, to encode a sequence of informational vectors it is necessary to:

- define a level Λ of the distortion correction,
- define a number r of check vectors using (2) and (3),

- obtain a code polynomial $c(x) = a(x)g(x) = c_{n-1}x^{n-1} + ... + c_1 x + c_0$ where $n = k + r$,
- assign each vector of codeword $c_{n-1}c_{n-2}...c_1c_0$ to a BC pattern and locate BC patterns in a shape of a rectangle matrix, i.e. produce a BC symbol.

5. Errors and erasures correction in the decoding process

While scanning, a barcode image processing of BC patterns occurs sequentially. At the moment of decoding, a decoder knows an overall number n of BC patterns in the BC symbol as well as a number r of control BC patterns.

If it is impossible to identify a current BC pattern with any of BC patterns of a barcode symbolism, this BC pattern is considered as erased. The location of the erased BC pattern (we will name it a locator) is known, because an order number of the BC pattern in the BC symbol is known. However, a value of the distorted pattern is unknown. Our task is to restore it. To do this, any predefined value should be assigned to the erased BC pattern; we assign zero value to the erased symbols.

The decoder task is to correct distortions of two types: (1) erasures for which locators are known and distortion values are unknown, and (2) errors for which neither locators nor distortion values are known.

Let us consider two different locator polynomials: a polynomial of erasure locators

$$\lambda(x) = \prod_{l=1}^{\delta}(1 - X_{i_l}x) \tag{8}$$

where X_{i_l} is erasure locators and δ is a number of erasures, and a polynomial of error locators

$$\sigma(x) = \prod_{j=1}^{\gamma}(1 - X_{p_j}x) \tag{9}$$

where X_{p_j} is error locators and γ is a number of errors.

The procedure of BC symbol decoding is the following.

Step 1. We store the numbers $i_1, i_2, ..., i_\delta$ of the digit places of a scanned word which contain erasures. Next, we need to check a ratio between δ and r. If $\delta > r$, a number of erasures exceeds correcting ability of error-correcting code and decoding process should be stopped due to inoperative distortions. If $\delta \leq r$, then erasure locators $X_{i_1} = \alpha^{i_1}, X_{i_2} = \alpha^{i_2}, ..., X_{i_\delta} = \alpha^{i_\delta}$ must be found. The digit places where the erasure is identified are assigned to zero.

Step 2. We find a polynomial (8) of erasure locators $\lambda(x)$ from the given values $X_{i_l}, l = 1, 2, ..., \delta$.

Step 3. We identify a maximum number χ of errors that decoder can correct after identifying δ erasures:

$$\chi =](r - \delta)/2[. \tag{10}$$

The actual number γ of errors must satisfy the condition $0 \le \gamma \le \chi$.

Step 4. We calculate components $S_1, S_2, ..., S_r$ of distortion syndrome:

$$S_w = c'(\alpha^w), w = 1, 2, ..., r \tag{11}$$

where $c'(x) = \sum_{i=0}^{n-1} c'_i x_i$ and c'_i is digit places of the scanned word. If all components of the syndrome are equal to zero, then there is no distortion in the scanned word.

Step 5. We construct a syndrome polynomial by using given components $S_1, S_2, ..., S_r$ of the syndrome:

$$S(x) = 1 + \sum_{w=1}^{r} S_w x^w. \tag{12}$$

Next, we search for a modified syndrome polynomial of distortions $V(x)$ from the ratio $S(x)\lambda(x) \equiv V(x) \bmod x^{r+1}$ that is:

$$(1 + \sum_{w=1}^{r} S_w x^w)\lambda(x) = 1 + \sum_{w=1}^{r} V_w x^w \bmod x^{r+1} \tag{13}$$

where $V(x) = 1 + \sum_{w=1}^{r} V_w x^w$ and $V_1, V_2, ..., V_r$ are components of the modified distortion syndrome. The polynomial of erasure locators $\lambda(x)$ derives from the ratio (8).

Step 6. We find an error syndrome. To do this, we separate a subsequence V_j, $V_{j+1}, ..., V_r$ from the sequence $V_1, V_2, ..., V_r$, where $j = r - 2\chi + 1$, represent it as $D_1, D_2, ..., D_{2\chi}$ where $D_\varepsilon = V_{r-2\chi+\varepsilon}$ and $\varepsilon = 1, 2, ..., 2\chi$.

Then $D_1, D_2, ..., D_{2\chi}$ are components of the error syndrome.

Step 7. We find a polynomial of error locators from (9). The coefficients σ_j of a polynomial $\sigma(x) = 1 + \sum_{j=1}^{\gamma} \sigma_j x^j$ can be calculated based on the components D_ε of an error syndrome.

Since a number of erasures δ is known, we consider that a number γ of errors is equal to χ, i.e. $\gamma = \chi$.

We compose a matrix equation:

$$\begin{Vmatrix} D_1 & D_2 & \cdots & D_\gamma \\ D_2 & D_3 & \cdots & D_{\gamma+1} \\ \cdots & \cdots & \cdots & \cdots \\ D_\gamma & D_{\gamma+1} & \cdots & D_{2\gamma-1} \end{Vmatrix} \times \begin{Vmatrix} \sigma_\gamma \\ \sigma_{\gamma-1} \\ \cdots \\ \sigma_1 \end{Vmatrix} = - \begin{Vmatrix} D_{\gamma+1} \\ D_{\gamma+2} \\ \cdots \\ D_{2\gamma} \end{Vmatrix}. \tag{14}$$

Next, we find a determinant of matrix M:

$$M = \begin{Vmatrix} D_1 & \cdots & D_\gamma \\ \cdots & \cdots & \cdots \\ D_\gamma & \cdots & D_{2\gamma-1} \end{Vmatrix}.$$

If $\det M \neq 0$, it means that the chosen γ is correct.

Then we can solve the matrix equation (14):

$$\begin{Vmatrix} \sigma_\gamma \\ \cdots \\ \sigma_1 \end{Vmatrix} = M^{-1} \begin{Vmatrix} -D_{\gamma+1} \\ \cdots \\ -D_{2\gamma} \end{Vmatrix}. \tag{15}$$

If $\det M = 0$, the value γ must be decreased by one and the procedure must be repeated. The reduction of γ should be performed until we get a determinant of matrix M other than zero. After that M^{-1} is found and coefficients σ_j are calculated.

Step 8. We solve an equation $\sigma(x) = 0$ in the field $GF(P_\Omega)$ and find roots $x_1, x_2, ..., x_\gamma$. Then the error locators $X_{p_1}, X_{p_2}, ..., X_{p_\gamma}$ are equal to $X_{p_1} = x_1^{-1}$, $X_{p_2} = x_2^{-1}$, ..., $X_{p_\gamma} = x_\gamma^{-1}$.

Step 9. We find a polynomial of distortions $Q(x)$ of power r from a ratio $V(x)\sigma(x) \equiv Q(x) \bmod x^{r+1}$ or $(1 + \sum_{w=1}^{r} V_w x^w)\sigma(x) \equiv Q(x) \bmod x^{r+1}$.

Step 10. We compose a joint set X of distortion locators which consists of the error locators and the erasure locators: $X = \{X_1, X_2, ...X_{\delta+\gamma}\}$.

Step 11. We calculate values E_{beta} of distortions (both errors and erasures):

$$E_\beta = \frac{Q(X_\beta^{-1})}{\prod_{\beta=1,\ j=1,\ \beta\neq j}^{\delta+\gamma}(1 - X_\beta^{-1}X_j)}. \tag{16}$$

Step 12. We correct distortions. For each pair $\langle X_\beta, E_\beta\rangle$, where $\beta = 1, 2, ..., \delta + \gamma$, X_β must be rewrited as $X_\beta = \alpha^f$ and f must be found. In the digit place f of the scanned word the correction must be conducted: $c_f = (c'_f - E_\gamma) \bmod m_s(x)$.

If X_β are erasure locators, then E_β are values of the difference between the chosen values of erased symbols with locators X_β and the true symbols. If X_β are error locators, then E_β are values of the difference between the scanned symbols with locators X_β and the true symbols.

The result of this procedure is the initial encoded data recovered from the damaged BC symbol.

6. Experiments

Let us use the developed method for the following example. Let the scanned word be equal to 403 0 179 100 597 445 526 506 686. Then we represent it as follows:

$$\begin{array}{rccccccccc} X_i: & \alpha^8 & \alpha^7 & \alpha^6 & \alpha^5 & \alpha^4 & \alpha^3 & \alpha^2 & \alpha^1 & \alpha^0 \\ c': & 403 & \mathbf{0} & 179 & \underline{100} & 597 & 445 & 526 & 506 & 686 \end{array}$$

It corresponds to the polynomial:

$$c'(x) = 403x^8 + 179x^6 + 100x^5 + 597x^4 + 445x^3 + 526x^2 + 506x + 686.$$

Let us assume that there is an erasure in the digit place 7, where a decoder has replaced zero (shown as bold), and there is an error in the digit place 5 (shown as underlined).

The decoding process consists of the following stages:

1. We store the numbers $i_1, i_2, ..., i_\delta$ of digit places, which contain erasures, as well as the number δ of erasures. In the given example, $i_1 = 7$ and $\delta = 1$.

We check whether $\delta \leq r$, i.e. we define if the error-correcting code is able to correct this number of erasures. Since $r = 3$, the correction is possible.

2. We find erasure locators by using the known values of digit place numbers where erasures occurred: $X_{i_1} = \alpha^7 = 21$.

Next, we can find a polynomial of erasure locators using (8): $\lambda(x) = 1 - 21x = 1 - \alpha^7 x = 1 - (2\alpha^2 + \alpha)x = 1 + (\alpha^2 + 2\alpha)x = 1 + 15x = 1 + \alpha^{371}x$.

3. We calculate the maximum number χ of errors a decoder can correct after the detection of δ erasures from (10):

$$\chi =](r - \delta)/2[=](3 - 1)/2[= 1.$$

Thus, it is possible to correct one error.

4. We calculate the components $S_1, S_2, ..., S_r$ of distortion syndrome by using (11). We consider $r = 3$:

$$S_1 = c'(\alpha^1) = 403\alpha^8 + 179\alpha^6 + 100\alpha^5 + 597\alpha^4 + 445\alpha^3 + 526\alpha^2 + 506\alpha + 686 = \alpha^{342} \cdot \alpha^8 + \alpha^{303} \cdot \alpha^6 + \alpha^{158} \cdot \alpha^5 + \alpha^{47} \cdot \alpha^4 + \alpha^{672} \cdot \alpha^3 + \alpha^{339} \cdot \alpha^2 + \alpha^{346} \cdot \alpha + \alpha^{439} = \alpha^{350} + \alpha^{309} + \alpha^{163} + \alpha^{51} + \alpha^{675} + \alpha^{341} + \alpha^{347} + \alpha^{439} = \alpha^{489} = 291.$$

Table 3 shows digit-by-digit addition of the coefficients modulo 3.

Table 3: Digit-by-digit addition of the coefficients modulo 3

α^{350}	0 0 2 0 1 2
α^{309}	1 2 2 2 0 2
α^{163}	1 2 1 1 2 0
α^{51}	1 1 1 0 1 0
α^{675}	1 1 0 2 1 1
α^{341}	1 1 1 2 2 0
α^{347}	0 0 2 0 0 2
α^{439}	2 2 1 1 0 2
$\alpha^{489} = 291$	1 0 1 2 1 0

$$S_2 = c'(\alpha^2) = 403\alpha^{16} + 179\alpha^{12} + 100\alpha^{10} + 597\alpha^8 + 526\alpha^4 + 506\alpha^2 + 686 = \alpha^{342} \cdot \alpha^{16} + \alpha^{303} \cdot \alpha^{12} + \alpha^{158} \cdot \alpha^{10} + \alpha^{47} \cdot \alpha^8 + \alpha^{672} \cdot \alpha^6 + \alpha^{339} \cdot \alpha^4 + \alpha^{346} \cdot \alpha^2 + \alpha^{439} = \alpha^{565} = 94$$

$$S_3 = c'(\alpha^3) = 403\alpha^{24} + 179\alpha^{18} + 100\alpha^{15} + 597\alpha^{12} + 445\alpha^9 + 526\alpha^6 + 506\alpha^3 + 686 = \alpha^{247} = 86$$

5. We compose a syndrome polynomial from (12):

$$S(x) = 1 + 291x + 94x^2 + 86x^3 \text{ or } S(x) = 1 + \alpha^{489}x + \alpha^{565}x^2 + \alpha^{247}x^3.$$

We calculate modified syndrome distortion polynomial $V(x)$ from (13):

$$V(x) = S(x)\lambda(x) \bmod x^{r+1} = (1 + \alpha^{489}x + \alpha^{565}x^2 + \alpha^{247}x^3) \cdot (1 + 15x) \bmod x^4 = 1 + \alpha^{446}x + \alpha^{59}x^2 + \alpha^{64}x^3 = 1 + 270x + 477x^2 + 657x^3.$$

6. Now we can find an error syndrome. We extract a subsequence $V_j, V_{j+1}, ..., V_r$ where $j = r - 2\chi + 1 = 3 - 2 + 1 = 2$, and represent it as $D_1, D_2, ..., D_{2\chi}$, i.e. components of the error syndrome. Since $D_\varepsilon = V_{r-2\chi+\varepsilon}$, $\varepsilon = 1, 2, ..., 2\chi$ and in our case, $\chi = 1$, $\varepsilon \in \{1; 2\}$ we get the following components:

$$D_1 = V_{3-2\cdot1+1} = V_2 = 477 = \alpha^{59}, D_2 = V_{3-2\cdot1+2} = V_3 = 654 = \alpha^{64}.$$

7. We calculate a polynomial $\sigma(x)$ of the error locators. The coefficients σ_j of the polynomial are calculated based on the components D_ε of the error syndrome by using the matrix method.

Since the initial value is $\gamma = \chi = 1$, a matrix M consists of the only element:

$$M = \|D_1\| = \|\alpha^{59}\| = \|477\|.$$

As $\det M = 477 \neq 0$, the scanned word contains one error.

We can form a matrix equality (15):

$$\|\sigma_1\| = M^{-1} \times \| - D_{2\gamma}\|.$$

Now we can find M^{-1}:

$$M^{-1} = \|477\|^{-1} = \|\alpha^{59}\|^{-1} = \|\alpha^{-59}\| = \|\alpha^{669}\| = \|205\|.$$

The matrix equation becomes the following:

$$\|\sigma_1\| = \|\alpha^{669}\| \times \| - \alpha^{64}\| = \|\alpha^{669+64-728}\| = \| - \alpha^5\| = \|2\alpha^5\| = \|\alpha^{364}\| \times \|\alpha^5\| = \|\alpha^{369}\| = \|484\|.$$

Thus, the polynomial of the error locators is:

$$\sigma(x) = 1 + \alpha^{369}x.$$

8. We solve the equation $\sigma(x) = 0$ in the field $GF(3^6)$:

$$1 + \alpha^{369}x = 0$$

wherefrom $x = \frac{-1}{\alpha^{369}} = \frac{2}{\alpha^{369}} = \frac{\alpha^{364}}{\alpha^{369}} = \alpha^{-5}$.

Then an error locator is equal to $X_P = x^{-1} = \alpha^{-5}$, which means that the error occurred in the digit place 5.

9. We find a polynomial $Q(x)$ of power r of distortion values:

$$Q(x) = V(x)\sigma(x) \bmod x^4 = (1 + \alpha^{446}x + \alpha^{59}x^2 + \alpha^{64}x^3) \cdot (1 + \alpha^{369}x) \bmod x^4 = 1 + \alpha^3 x + \alpha^{563}x^2.$$

10. We form a joint set X of distortion locators that contains the erasure locators and the error locators: $X = \{\alpha^5, \alpha^7\}$, $X_1 = \alpha^5$, $X_2 = \alpha^7$.

11. We calculate values E_β of distortions, both erasures and errors, by using (16):

$$E_1 = \frac{Q(X_1^{-1})}{1 - X_1^{-1}X_2}, \; E_2 = \frac{Q(X_2^{-1})}{1 - X_2^{-1}X_1},$$

$$E_1 = \frac{1 + \alpha^3 \cdot \alpha^{-5} + \alpha^{563} \cdot \alpha^{-10}}{1 - \alpha^{-5} \cdot \alpha^7} = \frac{1 + \alpha^{-2} + \alpha^{553}}{1 - \alpha^2} = \frac{1 + \alpha^{726} + \alpha^{553}}{1 + 2\alpha^2} = \frac{1 + \alpha^{726} + \alpha^{553}}{1 + \alpha^{364} \cdot \alpha^2} = \frac{\alpha^{49}}{\alpha^{243}} = \alpha^{-194} = \alpha^{534},$$

$$E_2 = \frac{1 + \alpha^3 \cdot \alpha^{-7} + \alpha^{563} \cdot \alpha^{-14}}{1 - \alpha^{-7} \cdot \alpha^5} = \frac{\alpha^9}{\alpha^{605}} = \alpha^{-596} = \alpha^{132}.$$

12. Now we can correct distortions as follows:

$$c_1 = c_1' - E_1 = 100 - \alpha^{534} = \alpha^{158} + 2\alpha^{534} = \alpha^{158} + \alpha^{364} \cdot \alpha^{534} = \alpha^{158} + \alpha^{170} = 010201 + 102002 = 112200 = \alpha^{118} = 396,$$

$$c_2 = c_2' - E_2 = 0 - \alpha^{132} = 2\alpha^{132} = \alpha^{496} = 697.$$

After the correction (the error has been corrected in the digit place 5 and the erasure has been corrected in the digit place 7) the word becomes 403 697 179 396 597 445 526 506 686.

To obtain the informational word from this word, the polynomial $c(x) = 403x^8 + 697x^7 + 179x^6 + 396x^5 + 597x^4 + 445x^3 + 526x^2 + 506x + 686$ which corresponds to the corrected word should be divided by the generator polynomial (5). After fulfilling this operation, we obtain the informational word 403 470 262 436 720 302 which does not contain any distortions.

7. Results and discussion

In the general case, encoding the input sequence of alphanumeric data with a error-correcting code is optional. However, the practice of using barcoding data shows that the barcode data should be noise-immune, i.e. the sequence of informational vectors should be encoded with a error-correcting code and only then be place on a carrier as a BC symbol.

Let us analyze capability of the error-correcting (n, k)-code which can correct distortions of two types, erasures and errors, and ensure noise immunity of a BC symbol if either a part or a whole barcode image is damaged.

From (2) and (3), we obtain:

$$r = 2^{\Lambda} - 1. \tag{17}$$

On the other hand, $2\gamma + \delta \leq r$.

Thus, a permissible number of damaged BC patterns which allows to process a BC symbol for a given Λ and, therefore, a known r, lies within the range from $(r+1)/2$ to r. The former means that we can correct $(r-1)/2$ errors and one erasure, while the latter means that we can correct only one erasure.

To calculate this number, we use the arithmetic mean of these values. Then averaged number F_c of damaged BC patterns in a BC symbol for a given r is equal to:

$$F_c =]((r + 1)/2 + r)/2[=](3r + 1)/4[.$$

For example, the permissible number of damaged BC patterns for which the correct processing a BC symbol is yet possible for a given $r = 7$ lies within the range 4-7 and, thus, the averaged permissible number of damaged BC patters is equal to 5.

If erasures are unlikely, it means that the permissible number of damaged BC patterns in a BC symbol for a given r is equal to $F_a = (r - 1)/2$, where F_a is a lower asymptotic value, i.e. a lower bound.

The permissible area of a BC symbol damage for which the barcode can be processed correctly is defined by the formula:

$$S_c(k, r) = \frac{F_c}{k + r} = \frac{](3r + 1)/4[}{k + r}, \tag{18}$$

which gives an averaged value and the formula:

$$S_a(k, r) = \frac{F_a}{k + r} = \frac{](r - 1)/2[}{k + r}, \tag{19}$$

which defines an asymptotic value, i.e. lower bound.

The ratio of r, a number of check vectors, to k, a number of informational vectors, i.e. the informational capacity of a BC symbol, is the redundancy $R = r/k$. Let us define S_c and S_a as functions of R. The formulas of S_c and S_a for large enough k and r can be represented as follows:

$$S_c(R) \approx \frac{]3r/4[}{k+r} = \frac{3}{4} \cdot \frac{R}{1+R},$$

$$S_a(R) = \frac{r}{2(k+r)} = \frac{1}{2} \cdot \frac{R}{1+R}.$$

The functions $S_c(R)$ and $S_a(R)$ show how the permissible area of a BC symbol area, for which the correct recognition of a barcode is still possible, depends on the redundancy R and a number of check vectors r (see Fig. 3).

For example, if a number of check vectors is the same as a number of informational vectors, i.e. $k = r$, which means the doubling of a BC symbol, then the permissible area of a barcode image damaging is in the range from 25% to 37%, where 25% is the lower asymptotic value (only errors occur when damaged) and 37% is the averaged value (apart from errors, erasures also can occur).

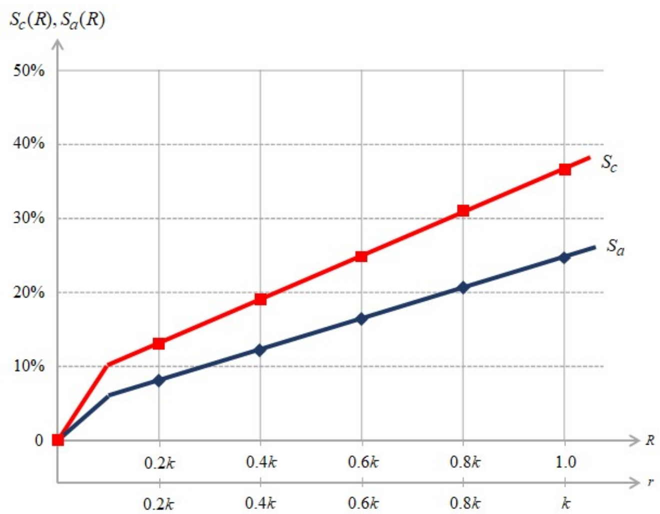

Figure 3: Dependence of the permissible area S of a BC symbol damaging on a redundancy coefficient R

If we define the maximum level of distortion correction from the ratio $2^{\Lambda} - 1 = (2/3)k$ (i.e. when a number r of check vectors is equal to two thirds of a number k of informational vectors and $r = 0.67k$), then the permissible area of a BC symbol damaging is of the range from 20% to 30%.

Thus, taking into account conditions where barcode object identification will be used, noise immunity of BC symbols can be ensured by choosing an appropriate number of check vectors when a barcode image is being created.

8. Conclusions

For reliable storing data as a three-color barcode label created by using BGW-Code technology, an error-correction code must be applied during data encoding in order to correct multiple distortions of multi-valued characters. The authors propose to use the Reed-Solomon code for this purpose.

The proposed method for correction of possible distortions in black-gray-white barcode labels provides a high level of noise immunity and reliability of data input into the computer system, which in turn expands the scope of information barcoding by using the proposed BGW-Code.

Further work may focus on improving noise-immune barcoding methods, as well as investigating the error-correction capability of multicolored (more than three-color) barcodes.

References

[1] *Guide to Barcode Types and Standards: 1D, 2D Barcode Symbologies, Requirements, and Standards-Issuing Entities*, Available at: https://www.camcode.com/asset-tags/guide-to-barcode-types-standards/.

[2] Pang, P., Wu, J., and Long, C., *CodeCube: A multi-layer color barcode for mobile social applications*, In: 29th Chinese Control and Decision Conference (CCDC), 2017.

[3] Bulan, O. e. a., *High capacity color barcodes using dot orientation and color separability*, Media Forensics and Security, Vol. 7254, 2009.

[4] Bagherinia, H. and R., M., *A theory of color barcodes*, In: IEEE Int. Conf. Comp. Vision Workshops, 2011.

[5] Barrus, J. and Wolff, G. J., *Embedding barcode data in an auxiliary field of an image file*, U.S. Patent No. US10/865,584; appdate 2004-06-09; pubdate 2005-12-15.

[6] Grillo, A. e. a., *High capacity colored two-dimensional codes*, In: Int. Multiconf. Computer Science and Information Technology, 2010.

[7] *High Capacity Color Barcodes (HCCB)*, Available at: http://research.microsoft.com/en-us/projects/hccb/.

[8] Bulan, O., Monga, V., and Sharma, G., *Method for encoding and decoding data in a color barcode pattern*, U.S. Patent No. US12/436,456; appdate 2009-05-06; pubdate 2010-11-11.

[9] *JAB Code*, Available at: https://github.com/jabcode/jabcode.

[10] Querini, M. e. a., *2D Color Barcodes for Mobile Phones*, International Journal of Computer Science and Applications, Vol. 8, No. 1, 2011.

[11] Cattrone, P., *Two-dimensional color barcode and method of generating and decoding the same*, U.S. Patent No. US11/444,288; appdate 2006-05-31; pubdate 2007-12-06.

[12] Parikh, D. and Jancke, G., *Localization and segmentation of a 2D high capacity color barcode*, In: IEEE Workshop on Applications of Computer Vision, 2008.

[13] Wang, G., Yang, Z., Chen, J., Wattanachote, K., and Liu, W., *Security Mechanism Improvement for 2D Barcodes using Error Correction with Random Segmentation*, In: 6th International Conference on Information Technology: IoT and Smart City, 2018.

[14] Kamijo, K., Kamijo, N., and G., Z., *Invisible barcode with optimized error correction*, In: 15th IEEE International Conference on Image Processing, 2008.

[15] Wesley Peterson, W. and Weldon, E. J., *Error-Correcting Codes*, 1961.

[16] Blahut, R. E., *Theory and practice of error control codes*, Addison-Wesley, 1983.

[17] Kasami, T., Tokura, N., Ivadari, E., and Inagaki, Y., *Theory of coding*, 1978.