Research on Two-Dimensional Bar Coding Technology of Pdf417

Li-Jun MAO\textsuperscript{1,a} and Wen-Hao ZHANG\textsuperscript{2,b}

\textsuperscript{1}Intelligent Science and Information Engineering College, Xi'an Peihua University, Xi'an, China
\textsuperscript{2}Shaanxi Chunjun Electric Power Equipment Co., Ltd.
\textsuperscript{a}35909250@qq.com, \textsuperscript{b}39459046@qq.com

Keywords: Pdf417; Two dimensional bar code; Code.

Abstract. This paper mainly introduces the structure and encoding principle of PDF417 two-dimensional barcode, and gives encoding algorithms such as data code-word encoding, error correction code-word encoding and symbol code-word encoding. Finally, the barcode pictures are generated through experiments.

Introduction

Bar code technology is a new data recognition technology developed on the basis of computer technology and information technology. With the rapid development of Internet technology and multimedia technology, one-dimensional barcode can’t meet the needs of describing various information of objects in practical application. This is because one-dimensional barcodes are easy to be limited by information capacity and depend on the database when they are used. Therefore, the birth of two-dimensional barcodes makes up for these shortcomings. The application of two-dimensional barcode technology has become a highlight in current data acquisition and information automatic identification. Especially in information security identification, the use of two-dimensional barcode technology is an effective anti-counterfeiting means adopted by all countries in the world. PDF417 code, which is the most representative of many two-dimensional bar codes, is widely used.

Introduction of PDF417 Two-dimensional Bar Code

PDF417 bar code is a row-and-row two-dimensional bar code invented by Dr. Yinjun Wang, a Chinese studying in the United States. It was officially published in 1991. It is a widely used and mature linear stacked two-dimensional bar code. PDF is called Portable Data File in English, which means "Portable Data File". Each symbol of the code is composed of four bars and four empty 17 modules, so it is called PDF417 barcode.

PDF 417 bar code symbol is a multi-line structure, which consists of the same number of characters in each line of data. Each line is aligned up and down. The number of rows ranges from 3 to 90 rows. Each line of PDF417 barcode symbol is structurally composed of seven parts: left blank area, start, left indicator, data area, right indicator, terminator and right blank area. The starters and terminators on the left and right sides are determined and will not be changed. Therefore, we focus on coding Left-Line indicators, data areas and Right-Line indicators.

PDF417 Barcode Coding

PDF417 barcode coding is divided into three parts: data codeword coding, error correction codeword coding and symbol codeword coding.

Data Code Coding Algorithms

PDF417 barcode uses three data compression modes to form character set. Three data compression modes are text compression mode (TC), byte compression mode (BC) and digital compression mode (NC). By using pattern locking/transfer codewords, data can be represented by multiple patterns in a PDF417 barcode symbol. Among them, mode locking is mainly to switch the
current mode to the specified target mode, which is always valid unless the next mode switching has started. Mode transfer refers to the switching of text compression mode (TC) to byte compression mode (BC), which is only valid for the first codeword after switching, and then the codeword returns to the current sub-mode of text compression mode (TC).

**Text Compression Mode.** Text compression mode can be divided into four sub-modes: Alpha, Lower Case, Mixed and Punctuation. In each seed mode, a set of characters with higher frequency in the file is selected as the character set. Each character in the sub-pattern corresponds to a numeric value ranging from 0 to 29, respectively. Thus, the two text characters can be coded into a 417 code word, which is equal to \((30 \cdot H) + L\), where \(H\) and \(L\) are the first and second text characters respectively. Generally, when converting from other modes to text compression mode, the default is to start with capital letter mode.

**Byte Compression Mode.** Byte sequence can be converted into codeword sequence by byte compression mode. The conversion from base 256 to base 900 is mainly used here. For byte compression mode, there are two mode locks (901, 924). If the number of compressed bytes is a multiple of 6, lock in 924 mode. According to the order from left to right, through the conversion from base 256 to base 900, 6 bytes can be converted into 5 codewords, such as 6 bytes A1, A2, A3, A4, A5, A6 can be represented as a codeword sequence 924, B1, B2, B3, B4, B5. The conversion method is as follows:

\[
A1 \times 256^5 + A2 \times 256^4 + A3 \times 256^3 + A4 \times 256^2 + A5 \times 256 + A6 = B1 \times 900^4 + B2 \times 900^3 + B3 \times 900^2 + B4 \times 900 + B5 \quad (1)
\]

If the number of compressed bytes is not a multiple of 6, it is locked in 901 mode. The first 6 data bytes are still converted into 5 codewords according to the above method, and the remaining 1-5 bytes after dividing by 6 correspond to one codeword respectively.

**Digital Compression Mode.** Digital compression mode refers to the conversion from base 10 to base 900. It can approximately represent three decimal digits in one code. Although the digital compression mode does not limit the length of words that need to be compressed, it is generally recommended to use it when the continuous digit bits are greater than or equal to 13, otherwise the text compression mode is adopted.

Digital compression mode coding can be divided into: 1. Divide the digital sequence into a group of 44 bits from left to right, the last group contains less than 44 bits; 2. For each group of digits, add a preamble "1" before its digital sequence, and then make a cardinal conversion from 10 to 900. Example: "100213298174002". Because the number sequence is less than 44 bits, it can only be divided into a group, preceded by a preamble 1, and then converted from 10 to 900.

\[
100213298174002 = 1 \times 900^5 + 624 \times 900^4 + 434 \times 900^3 + 632 \times 900^2 + 282 \times 900 + 202 
\]

So the coding sequence of 417 barcode is \((1, 624, 434, 632, 282, 202)\).

**Error Correcting Codeword Coding**

When the data codeword in the data stream is generated, the error-correcting codeword can be generated. Given the error-correcting level \(s\), the length of the error-correcting codeword \(M = 2S + 1\) is calculated by the error-correcting level. The algorithm of \(M\) error-correcting codewords mainly includes the following steps [1]:

**Step 1:** Establish a symbolic data polynomial. The formula is as follows:

\[
d(x) = d_{n-1} x^{n-1} + d_{n-2} x^{n-2} + \ldots + d_1 x + d_0 \quad (3)
\]

The coefficients in the polynomial are composed of the generated data codewords, \(n\) is the length of the data codewords, where \(D_{n-1}\) corresponds to the first data codeword, and so on, \(d_0\) corresponds to the last data codeword.

**Step 2:** Establish the generating polynomial of error-correcting codewords. The formula is as follows:

\[
g(x) = (x-3) (x-3^2) \ldots (x-3^m) = x^n + g_{m-1} x^{m-1} + \ldots + g_1 x + g_0 \quad (4)
\]

\(m\) is the number of error-correcting codewords in the polynomial.

**Step 3:** Error-correcting codeword calculation. For a given set of data codewords and selected error correction levels, error correction codewords \(c_0, \ldots, C_{m-1}\) is a symbolic data polynomial \(D(x)\) multiplied by \(x_m\), and then divided by the generating polynomial \(g(x)\). The coefficients of the remainder are complements over the finite field GF (929).
Symbol Coding

After the data codeword and error correction codeword are generated, the line indicator character is also calculated. The arrangement position of PDF417 barcode symbol is shown in Table 1.

<table>
<thead>
<tr>
<th>Initial character</th>
<th>( L_1 )</th>
<th>( d_{n-1} )</th>
<th>( d_{n-2} )</th>
<th>( d_{n-L} )</th>
<th>( R_1 )</th>
<th>( L_2 )</th>
<th>( d_0 )</th>
<th>( c_{m-1} )</th>
<th>( c_{m-2} )</th>
<th>( R_2 )</th>
<th>( L_{k-1} )</th>
<th>( d_0 )</th>
<th>( c_{m-1} )</th>
<th>( c_{m-2} )</th>
<th>( R_{k-1} )</th>
<th>( L_k )</th>
<th>( c_1 )</th>
<th>( c_0 )</th>
<th>( R_k )</th>
</tr>
</thead>
</table>

Where \( K \) denotes the total number of rows of the code, \( L \) denotes the number of columns in the data area of the code, and the starting and ending characters of the PDF417 bar code are fixed values. The algorithm for line indicator characters is determined by the following formula, where the left indicator is \( L_i \), \( R_i \):

\[
L_i = \begin{cases} 
30x + y & c_i = 0 \\
30x + z & c_i = 3 \\
30x + v & c_i = 6 
\end{cases}
R_i = \begin{cases} 
30x + v & c_i = 0 \\
30x + y & c_i = 3 \\
30x + z & c_i = 6 
\end{cases}
\]

In formula: \( i \) denotes line number, \( xi=\text{INT}[(i-1)/3] \) \( i=1, 2, \ldots, k, \) \( y=\text{INT}[(k-1)/3], z=s*3 + (k-1) \) mod 3, \( v=L-1, c_i=\text{(i-1) mod 3}*3 \) Cluster number of line \( i \).

Then the data codeword, error-correcting codeword and line indicator character are converted into barcode null number by corresponding cluster number.

Generating Bar Code Pictures

In this paper, two functions of drawing black bars and drawing blanks are used to generate barcode pictures. One is to call drawing black bars function in odd digits per line of generated symbolic codeword sequence, and the other is to call drawing blank functions in even digits. The barcode picture can be drawn according to the combination of bars and blanks. Because the barcode picture only has black and white color, the BMP format picture which generates monochrome bitmap can compress the storage space of the picture very well.

Bar Code Original Information: PDF417 Two-Dimensional Bar Code

The bar code generated by the software is shown in Fig. 1.

![Figure 1. Generates a Bar Code Picture.](image)

Conclusion

This paper mainly introduces the structure and coding principle of PDF417 two-dimensional bar code. According to its coding characteristics and principles, bar code pictures are generated by experiments. The experimental results show that the generated PDF417 two-dimensional bar code can be easily applied to real life.

Acknowledgment

This work was supported by the Natural Science Foundation of Xi’an Peihua university “Research on Dynamic Migration Algorithm of Open Flow Switch Based on Attractor Selection in Cloud
Environment” under Grant No. PHKT18062. It is also supported by The second batch of cooperative education projects of industry-university cooperation in 2018, "Research on the development of cooperative curriculum between private colleges and enterprises under the background of new subjects” under Grant No. 201802153008 and “Reform of Applied Course of Java Language Programming” under Grant No. 201802070035.

References


