An Implementation of a High Capacity 2D Barcode

Puchong Subpratatsavee[1] and Pramote Kuacharoen[2]

Department of Computer Science, Graduate School of Applied Statistics
National Institute of Development Administration
118 Serithai Rd. Bangkapi, Bangkok 10240 Thailand
puchong.sp@gmail.com[1], pramote@as.nida.ac.th[2]

Abstract. Two-dimensional barcode technology is becoming popular and useful for communication as well as transport data. However, the common two-dimensional barcodes can contain a very little amount of data. Although certain types of two-dimensional barcodes can contain more data, the size of the barcode image becomes too large. As a result, it is impractical to insert or attach the barcode to the document in the form of paper or print media. The barcode would occupy a large portion of the page. The space would be wasted and the information may not fit on the page.

This paper presents the design and the implementation of a high capacity two-dimensional barcode which shape and size are suitable for attaching to the document in the form of paper or print media. The proposed implementation reduces space usage compared to existing barcodes. The barcode can store both text and binary data. For space efficiency, the proposed implementation employs data compression. Therefore, the amount of data stored can vary widely depending on the content of the data. The maximum data size after compression is 24,400 bits.

Keywords: Two-dimensional barcode, 2D barcode, high capacity barcode

1 Introduction

Two-dimensional barcode or 2D barcode [1] is very popular for communication, public relations, and data transport. However, in general, the two-dimensional barcodes can contain a very little amount of data. Although the QR Code [2][3] can store more data than other two-dimensional barcodes, the size of the image of the QR Code becomes larger as the amount of data increases. As a result, it is impractical to insert or attach the QR Code which contains a higher amount of data to the document in the form of paper or print media since the QR Code occupies a significant amount of space.

This paper presents the design and the implementation of a two-dimensional barcode. The aim is to create a two-dimensional barcode which achieves a high capacity while it occupies a small area. The size of the barcode must be suitable for displaying on print media such as paper and poster. The proposed barcode must have a higher capacity than the existing two-dimensional barcodes.

This paper consists of five sections. The next section provides background information and related work that is relevant to the paper. Section 3 describes the design and implementation of the proposed barcode. Section 4 presents the experimental results. The last section, Section 5 concludes the paper.

2 Background and Related Work

This section provides background information related to this paper such as error correction function, data compression, and the existing two-dimensional barcodes.

2.1 Error Detection and Correction

Regardless of the design of the two-dimensional barcode, there will be errors. The image of the barcode may be damaged or unclear. In other words, the barcode obtained from scanning differs from the original. Therefore, it is essential to be able to detect and correct errors which may occur. There are many techniques for error detection and correction [4]. Zhang and Yuan [5] describe a multiple bits error correction method based on cyclic redundancy check (CRC).

The error correction which is widely used is Reed-Solomon code [6]. This is because it is easier to encode than other encoding techniques and the error correction code is small. Moreover, error detection and error correction can be performed quickly and efficiently. Reed-Solomon codes are prominently used in consumer electronics such as CDs DVDs, and Blue-ray Discs. The Reed-Solomon error correction is also used in two-dimensional barcodes such as PDF417 barcode, Datamatrix, Maxi Code, QR Code and Aztec Code. This allows correct reading even though a portion of the barcode is damaged.

2.2 Data Compression

The data compression [7] is a process of encoding data so that the number of bits used to store data is less than the original. Compression techniques are widely used in the data storage and communication. This allows storing and sending/receiving more data by using the same storage space or the same bandwidth for sending data comparing to uncompressed data. By compressing data, network transfer can be done more quickly. To obtain the original data, the compressed data is uncompressed.

2.3 Two-Dimension Barcodes (2D Barcodes)

Two-dimensional barcodes are geometric patterns in two dimensions. Two-dimensional barcodes can store more data than one-dimensional barcodes while using the same or smaller space since they can store data in both vertical and horizontal directions. Fig. 1 shows the characteristics of two-dimensional barcodes, namely; the QR Code and PDF417 Code compared with a typical one-dimensional barcode.

Fig. 1. Two-dimensional barcodes compared with a one-dimensional barcode

In addition, two-dimensional barcodes usually include error correction codes. If some parts of the two-dimensional barcode are deformed or damaged, it is possible to decode the barcode correctly.

Generally, two-dimensional barcodes contain black squares on a white background. Currently, two-dimensional barcodes are becoming very popular. It features both in terms of capacity to store data and reading speed. The two-dimensional barcodes that are very popular are QR Code, PDF417 barcode [8], Maxi Code, Aztec Code [9], and Data Matrix [10] [11]. Table 1 illustrates the properties of the two-dimensional barcodes.

Table 1. The characteristics and properties of two-dimensional barcodes.

	PDF417	Data Matrix	Maxi Code	QR Code	Aztec Code
Code type	Multi-low	Matrix	Matrix	Matrix	Matrix
Capacity (Characters)	1,850	2,355	93	4,296	3,067
Characteristic	High capacity	High capacity, small	High speed reader	High capacity, small, high speed reader	High capacity
Applications	Office	Plant, medical industry	Industrial products import and export	All industries	Aviation and transport industries
Standard	AIMI, ISO	AIMI, ISO	AIMI, ISO	AIMI, ISO, JIS	AIMI

From Table 1, different types of two-dimensional barcodes are created to serve different purposes. When compared with other barcodes, the QR Code has a high capacity while maintaining a small size and high reading speed. For these reasons, QR Code is used in public relations, communications, and applications for data storage.

3 Design and Implementation

This section presents the design and implementation of a high capacity two-dimensional barcode.

3.1 Design of a High Capacity Two-Dimensional Barcode

The objective of the design is to create a two-dimension barcode which has high capacity and occupies a small space. Furthermore, the size of the two-dimensional barcode should be suitable for appending to the document which will be printed on paper. Table 1 shows that QR Code has desirable characteristics of high capacity and small size. However, the QR Code is square. The space will be wasted unless the text is formatted to wrap around the barcode. On the other hand, PDF417 barcode is rectangular but has lower capacity.

Fig. 2 illustrates an example of the design of a High Capacity Two-Dimensional (HC2D) barcode. Similar to PDF417 barcode, the HC2D barcode is rectangular so space will not be wasted and there is no need to wrap text around the barcode.

Fig. 2. An Example of a HC2D barcode

The HC2D barcode consists of a vertical line on the left of the barcode and a horizontal line on the bottom. The top of the vertical line is a detection point. Another detection point is where the vertical line and horizontal line meet. The dash line on the top is used for sampling column widths. The vertical line on the right is the stop indicator. Finally, the data area is configured as a 63x448 matrix which can store black and white pixels up to 28,224 pixels. The characteristics of the HC2D barcode are shown in Fig. 3.

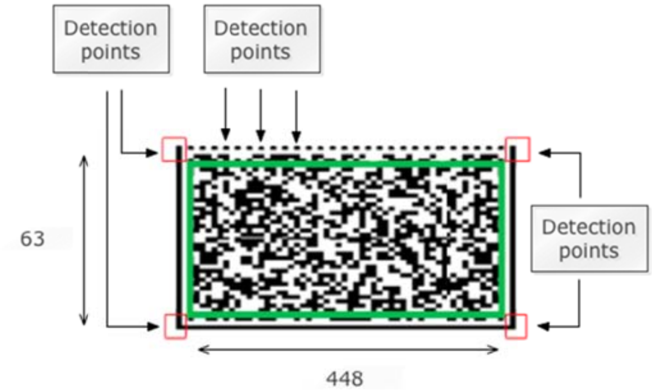

Fig. 3. The characteristics of HC2D barcode

3.2 HC2D Header Format

The header consists of the version of the barcode, data type, compression option, checksum, and the length of the data as illustrated in Fig. 4.

Fig. 4. The HC2D barcode data format

The description and the size of each field are described in Table 2.

Table 2. HC2D field description and size.

	Description	Size (bits)
Version	Version of the HC2D barcode	4
Data Type (DT)	The type of the data in the barcode	2
Compression Option (Z)	Compression option	2
Checksum	Header checksum	8
Length	The length of the data	16
Data	Data Stored in the barcode	Max. 24,400

The current version is 1. The data type field stores the type of the data in the barcode. The data type can be numeric data (00), ASCII characters (01), or binary (10). The compression option field indicates whether or not the contained data is compressed. The length field indicates the size of the data. If the types of the data are numeric data or ASCII characters, the length field indicates the number of digits or characters. Otherwise, it indicates the size in bytes. The checksum field is the header checksum which provides error detection capability. The checksum algorithm is CRC-8-CCITT. The length of the header is four bytes.

3.3 Recovering Data from a Damaged HC2D Barcode

The HC2D barcode contains error correcting code. Therefore, the correct reading of the barcode can still be achieved even though the barcode is damaged. The HC2D barcode uses Reed-Solomon error correcting code which is capable of recovering the data from a damaged barcode up to 32%. Fig. 5 shows a damaged barcode which can be read correctly.

The current implementation allows two levels of error correction. At the low level, approximately 7% of the code words can be restored. At the high level, approximately 32% of the code words can be corrected.

Fig. 5. A damaged HC2D barcode, but still readable

3.4 Capacity of HC2D Barcode

As aforementioned, the HC2D barcode can store black and white pixels up to 28,224 pixels which consist of data and Reed-Solomon error correcting code. Therefore, the capacity is directly related to the level of error correction. For the low level option, every pixel group consists of 109-byte data and 17-byte error correction code. This allows the HC2D barcode to store data up to 24,400 bits. For the high level option, every pixel group consists of 25-byte data and 17-byte error correction code. The HC2D barcode can store data up to 5,600 bits.

The HC2D barcode also employs data compression so that more data can be stored in the barcode. The amount of compression varies with the type of data, but a typical text such as English and source code will be reduced by 60 to 70 percent [12]. This allows the HC2D barcode to have higher capacity than fixed encoding schemes.

The capacity values provided previously are for data which are uncompressible or are not compressed when the compression option is turned off. These values are the minimum capacities for the HC2D barcode.

Numeric data which consist of 0-9 are encoded as 3-bit or 4-bit values depending on the digit. ASCII characters are encoded as 7-bit values. Unicode characters and other data are treated as binary data. Numeric character codes are illustrated in Fig. 6.

Characters	Binary Code
0	000
1	001
2	010
3	011
4	100
5	101

Characters	Binary Code
6	1100
7	1101
8	1110
9	1111

Fig. 6. Numeric character codes

Table 3 shows the capacity of the HC2D barcode for uncompressed data.

Table 3. The capacity of the HC2D barcode for uncompressed data.

	7% Error Correction	32% Error Correction
Numbers	6,100-8,133 characters	1,396-1,861 characters
ASCII Characters	3,485 characters	797 characters
Binary	24,400 bits	5,584 bits

3.5 Encoding and Decoding of the HC2D Barcode

As illustrated in Fig. 7, the process for creating a HC2D barcode involves obtaining data to be stored in the barcode, compressing the data if the compression option is selected, generating error correcting code using the Reed-Solomon algorithm, and generating the barcode.

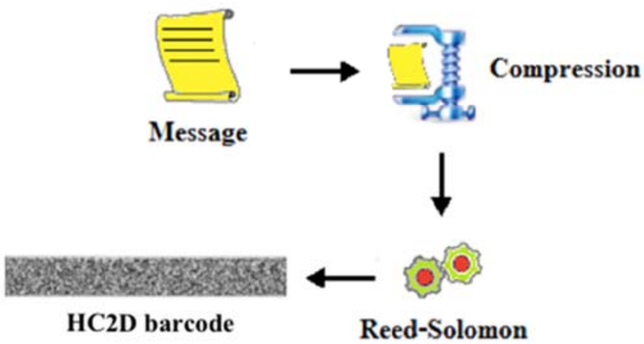

Fig. 7. The process for converting data into a HC2D barcode

For reading the HC2D barcode, the bit representation of the barcode is obtained by scanning the image of the barcode and converting the image. After obtaining the barcode which may be damaged, the Reed-Solomon error correction is performed. If the original data was previously compressed, the resulting data is then uncompressed. If there is an error in correcting the code, data decompression will generate an error. Otherwise, the original data will be obtained. The process for reading the HC2D barcode is shown in Fig. 8.

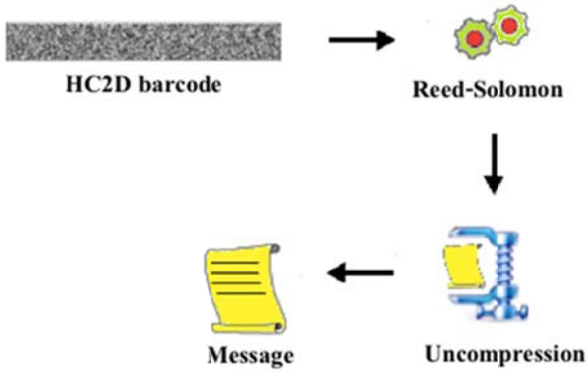

Fig. 8. The process for decoding data from a HC2D barcode

In order to detect the error correction level, the algorithm first processes the first pixel group as the low error correction level. If the header is obtained correctly, the error correction is at the low level. Otherwise, the algorithm processes the first pixel group at the high error correction level. The header verification is done using the checksum field. If the checksum value matched the one calculated, the assumed error correction level is accurate.

3.6 Applications of the HC2D Barcode

The HC2D barcode can be used in applications which require storing larger data than the available two-dimensional barcodes. For instance, the HC2D barcode can be used in paper-based document authentication [2]. Both the text information and the digital signature can be stored in the HC2D barcode without reducing the amount of text that can fit on the page.

4 Experimental Results

To verify our design, we implemented the HC2D barcode in Java programming language. The results are compared with QR Code and PDF417 barcode as described below.

4.1 Capacity comparison

Table 4 shows the capacity comparison between PDF417 barcode, QR Code and HC2D barcode. All types of barcodes are set at the minimum error correction level. Both PDF417 barcode and QR Code provide maximum capacity. However, HC2D barcode provides the minimum figures, which means that the HC2D could store more data if data compression is enabled.

Table 4. Capacity Comparison

	PDF417	QR Code	HC2D Barcode
Numbers	Max. 2,710	Max. 7,089	Min. 6,100-8,133
ASCII Characters	Max. 1,859	Max. 4,296	Min. 3,485
Binary	Max. 8,864 bits	Max. 23,624 bits	Min. 24,400 bits

For numeric data, PDF417 barcode and QR Code can store up to 2,710 and 7,089 characters, respectively. HC2D barcode encodes numeric data 0-5 using three bits each and 6-9 using four bits each. If the data contains only numbers 0-5, the barcode can store up to 8,138 characters. On the other hand, if the data contains only numbers 6-9, the barcode can store up to 6,103 characters. The given values are capacity for uncompressed data.

For alphanumeric data, PDF417 barcode can store up to 1,859 characters and QR Code can store up to 4,296 characters. QR Code only includes 45 characters, namely, 0-9, A-Z, nine special characters. HC2D barcode can store up to 3,488 characters which includes all ASCII characters.

For binary data, PDF417 barcode and QR Code can store up to 8,864 bits and 23,624 bits, respectively. HC2D barcode allows uncompressed binary data up to 24,400 bits.

Table 5 shows experimental results for the HC2D barcode with the compression option. The numeric data is randomly generated. The ASCII text is from this paper while Unicode text is from a random article in Thai. Finally, binary data is also randomly generated. As expected, random data cannot be efficiently compressed whereas text data can be compressed by more than 50%. Since PDF417 and QR Code do not have a compression option, they are not included in the experiment.

Table 5. The experimental capacity of the HC2D barcode with the compression option.

	7% Error Correction	32% Error Correction
Numbers	7,250 characters	4,700 characters
ASCII Characters	10,100 characters	6,000 characters
Unicode Characters	3,800 characters	2,350 characters
Binary	24,416 bits	16,800 bits

4.2 Barcode size comparison

The size of the barcode is very important when including a barcode to a document. A larger barcode size means fewer contents can be printed. The size of the barcode is usually proportional to the amount of data stored in the barcode. Fig. 9 shows the barcode size comparison and the amount of text that can be included so that both text and barcode can be printed on the same page. HC2D barcode can include more text than other barcode since the size of the barcode is smaller.

Note that the text which is used to generate barcodes is from this paper and the page size is A4.

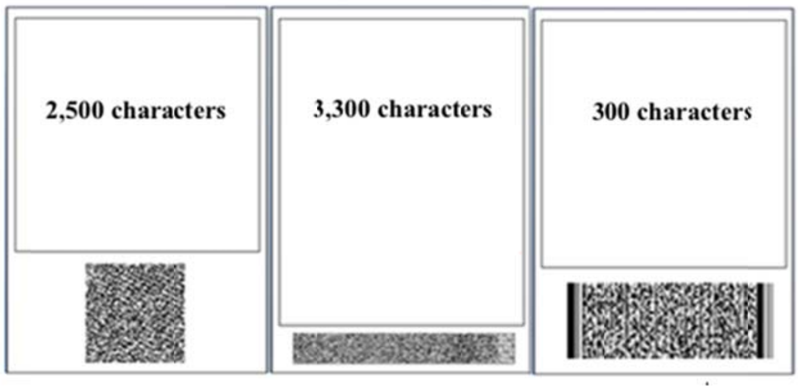

Fig. 9. Barcode size comparison between QR Code, HC2D barcode, and PDF147 barcodes

5 Conclusion

This paper presents an implementation of a high capacity two-dimensional barcode which has a greater capacity than existing 2D barcodes. Moreover, the shape of HD2D barcode is suitable for use with paper documents or print media. HC2D barcode can store uncompressed binary data up to 24,400 bits. However, with the use of data compression, more data may be stored in the HC2D barcode. Therefore, the maximum data size after compression is 24,400 bits.

References

1. Gao, J.Z., Prakash, L., Jagatesan, R.: Understanding 2D-Barcode Technology and Applications in M Commerce–Design and Implementation of a 2D Barcode Processing Solution. In: 31st Annual International Conference on Computer Software and Applications, pp.49-56, vol. 2, (2007)
2. Warasart, M., Kuacharoen, P.: Paper-Based Document Authentication Using Digital Signature and QR Code. In: Juan S.: 4th International Conference on Computer Engineering and Technology. International Proceedings of Computer Science and Information Technology, vol. 40, pp. 94-98 (2012)

3. QR Code, http://www.denso-wave.com/qrcode/
4. Singh, J., Singh, J.: A Comparative Study of Error Detection and Correction Coding Techniques. In: 2nd International Conference on Advanced Computing and Communication Technologies, pp. 187-189 (2012)
5. Zhang, Y., Yuan, Q.: A Multiple Bits Error Correction Method Based on Cyclic Redundancy Check Codes. In. 9th International Conference on Signal Processing, pp. 1808-1810 (2008)
6. Mamidi, S. et al.: Instruction Set Extensions for Reed-Solomon Encoding and Decoding. In: 16th IEEE International Conference on Application-Specific Systems, Architecture Processors, pp. 364-369 (2005)
7. Islam, M.R., Ahsan Rajon, S.A.: An Enhanced for Lossless Compression of Short Text for Resource Constrained Devices. In: 14th International Conference on Computer and Information Technology, pp. 292-297 (2011)
8. Rong, C. et al.: Coding Principle and Implementation of Two-Dimensional PDF417 Bar code. In: 6th IEEE Conference on Industrial Electronics and Applications, pp. 466-468 (2011)
9. Ke, H., Zhang, G.: An Algorithm Correcting Flex Distortion of Aztec Code. In: 2nd IEEE International Conference on Information Management and Engineering, pp. 457-460 (2010)
10. Biao, L. (2007), A DataMatrix-based mutant code design and recognition method research. In: Proceedings of the 4th international conference on image and graphics, pp. 570-574, 2007
11. Data Matrix, http://en.wikipedia.org/wiki/Data_Matrix
12. GNU Gzip, http://www.gnu.org/software/gzip/manual/gzip.html