Title:
An Evaluation Of 2D Barcodes In Document Processing Applications

Word Count:
1363

Summary:
Introduction

Many documents that are to be electronically processed contain barcodes to
encode important information that is extracted via barcode decode software.
There are a number of issues that should be considered when choosing a barcode
symbology. The largest distinguishing characteristic to be considered is whether
a linear (1D) or 2 Dimensional (2D) symbology is to be employed. 1D symbologies,
as the name implies, typically consist of width modulated bars and space...

Keywords:
2d,barcodes,document processing,bar code,data matrix,qr code,micro qr code

Article Body:
Introduction

Many documents that are to be electronically processed contain barcodes to
encode important information that is extracted via barcode decode software.
There are a number of issues that should be considered when choosing a barcode
symbology. The largest distinguishing characteristic to be considered is whether
a linear (1D) or 2 Dimensional (2D) symbology is to be employed. 1D symbologies,
as the name implies, typically consist of width modulated bars and spaces that
encode the user information. There is no information contained in the vertical
dimension of a 1D symbol. 2D symbologies encode information in both dimensions
of the symbol and as a result, have much higher data density. 2D symbols
typically use a regular grid of possible cell positions, where a cell is either
black or white.

This article will focus on the use of 2D symbols in document processing
applications because of the significant data density advantage of 2D symbols
over 1D symbols. In particular, we will compare the relative merits of three
popular public domain 2D symbologies: Data Matrix, QR Code and Micro QR Code.
Following a brief overview of each symbology, we will compare them based on
their data density, error correction, and relative processing speed.

Data Matrix

Data Matrix symbols use a regular array of square cells ranging in size from a 10 by 10 grid up to a 144 by 144 grid. A 1 cell quiet zone is required around the entire symbol. In addition, rectangular sizes are also available. Each symbol consists of a fixed "L" pattern that is used for finding along with a clock track along the opposite sides of the "L". In addition, there are internal clock tracks for larger Data Matrix. These fixed locations do not encode any information. They are present to identify the symbol as a Data Matrix and to aid the decode software. The remaining grid locations contain either a black or white squares depending on the information to be encoded.

QR Code

QR Code symbols also employ a regular array of square cells ranging in size from a 21 by 21 grid up to a 177 by 177 grid. A 4 cell quiet zone is required around the entire symbol. To aid finding, QR Code symbols contain 3 finder patterns at 3 of the 4 corners. In addition, there are internal alignment patterns, clock patterns, as well as format information on larger symbols that gives the size of the code.

For data applications that require smaller amounts of data, there is a derivative version of QR Code called Micro QR Code which can encode up to 35 numeric digits in less space than a corresponding QR Code. It has 4 different square sizes: 11 by 11, 13 by 13, 15 by 15 and 17 x 17. Each size requires a 2 cell quiet zone around the entire symbol. It contains only 1 finder pattern, with limited clock pattern and format information.

Data Density and Error Correction

Data Matrix has a clear data density advantage over QR Code. This is especially true for smaller amounts of user data. This is due to the fact that it has fewer fixed cell locations. It does not devote as much space for finder patterns, and contains no format information. Micro QR Code was designed to address the data density issue and is comparable in size to the Data Matrix for this data content.

All 3 types of symbols use Reed Solomon error correction to detect and correct errors due to symbol damage or imaging issues. The number of detectable and correctable errors is determined by the number of extra error correction codewords included in the symbol that are above and beyond the codewords used to encode the data.

The data capacity of a given size symbol is a function of the amount of error

correction overhead as well as the data itself. Data Matrix uses a fixed level of error correction that is not selectable by the user. The percentage of error correction codeword overhead ranges from 62.5% for the smallest symbol down to 28% for larger symbols. By contrast, QR Code has 4 different levels of error correction that allow an approximate recovery capacity of 7%, 15%, 25% or 30%. Micro QR Code varies the choices of the amount of error correction for each of the 4 allowable sizes. The smallest only allows error detection, while the largest allows up to 25% recovery capacity.

The amount and type of user data will dictate the size of the symbol that is necessary. In addition, for QR Code and Micro QR Code, the amount of error correction used will factor into the size as well. The table below summarizes the relative size and error correction capacities of the 3 symbols shown above.

Symbology -- Relative Size (with Quiet Zone) / Error Correction Overhead (%) / Maximum Correctable Errors

Data Matrix -- 1.00 / 58.3 / 3

QR Code -- 3.70 / 65.3 / 8

Micro QR Code -- 1.33 / 50.0 / 1

The choice of the amount of error correction used in QR Code and Micro QR Code is application dependent. In situations where size is an issue, one may be tempted to reduce the amount of error correction overhead. This may reduce the overall read rate of the symbol if the barcode may be damaged or if the imaging environment makes it more difficult to get "ideal" images. Barcodes on soft packages that curve the symbol, as well as shiny tape over the symbol that can cause specular reflection back to the camera are examples of how codes may be damaged. In general, if space permits, for optimum read rates, one should normally choose the maximum allowable error correction capacity.

Relative Processing Speed

In real time applications where the time to decode an image is important, one must also compare the symbologies on how quickly they can be decoded. The most time consuming part of decoding a barcode within a large and busy image is generally finding the symbol. The more unique the finder pattern within a barcode symbol, the easier it is to locate within a busy image. This reduces processing time. Conversely, if a barcode symbology does not provide a unique finder pattern, more time will be spent looking for it.

QR Code and Micro QR code have a significant advantage over Data Matrix because

of the unique finder patterns within the symbols. QR Code is the best of the 3 choices because it includes 3 finder patterns, each being able to be used to find the symbol. Data Matrix has the "L" finder pattern and fixed clock lines. Unfortunately, these are not terribly unique patterns with forms where many areas of text are surrounded by boxes. In addition, both QR (Version 7 and above) and Micro QR Codes have format information within the symbol to let you know the size of the symbol and to confirm you are on a real symbol. Data Matrix does not contain explicit format data, providing only a clock track on the opposite sides of the symbol from the "L" corner.

A busy form was scanned at 200 DPI, and a single instance of the 3 barcode symbols was added to the image with each symbol using 25 mil cells. Then in 3 separate passes, Volo™, a barcode decode software toolkit from Omniplanar®, was used to decode each symbol. In each pass, only one symbology type was enabled. The table below summarizes how long it took Volo to issue the decode result and completely finish processing the image. Both QR Code and Micro QR decoding were 3 to 4 times faster than Data Matrix decoding. This is almost entirely due to the good finder pattern in the QR and Micro QR symbols.

Symbology -- Issue Time (msecs) / Total Time (msecs)
Data Matrix -- 30.8 / 74.5
QR Code -- 7.2 / 23.4
Micro QR Code -- 7.6 / 21.9

Summary

When deciding what 2D symbology type to use in document applications, one must consider data density, error correction and processing time. In applications where the size of the symbol must be kept to a minimum, both Data Matrix and Micro QR code are good choices. When processing speed is of primary importance, QR Code and Micro QR Code are both better choices than Data Matrix given their good finder patterns. In applications when both symbol size and processing speed are important, Micro QR Code is the best choice. However the largest possible Micro QR Code can only store 35 numeric digits with the minimum error correction (maximum of 3 errors). At the maximum error correction level, the data capacity drops to 21 numeric digits (maximum of 7 errors).