Note 30	December 31, 2004

JPEG Image Compression Effects

Purpose: The compression methods applied to images by scanners or by imaging procedures can adversely affect their barcode readability. This note discusses the issues and suggests approaches for minimizing the problem.

JPEG Compression

There are several popular image compression methods: JPEG, GIF, TIFF and several less well known ones. JPEG compression is intended primarily for color and grayscale photographs. It has become increasingly popular with introduction of digital cameras, color scanners and the need to move compressed images from the web or via email. JPEG compression significantly reduces the image size and takes advantage of the human eye's tendency to gloss over fine image details.

JPEG employs a lossy compression scheme, which means that **each** time it is used, that a little of the information, some of the details, of the image are smeared out and are lost. The compression scheme essentially averages the color value of neighboring blocks of pixels. Consequently, JPEG losses occur in the fine detail, the edges and other sharp features of the image.

The amount of loss is governed by a parameter known as the Quality Factor, a dimensionless parameter whose attributes are not defined by any technical standard. Each vendor picks their own scale to use for this parameter. Some vendors use adjectives, such as High, Medium and Low. Other vendors use a misleading percentage scale, from 0 to 100. Picking 100 does not always preserve 100% of the information in the image, nor does a 0 setting destroy all the information. The value is merely an arbitrary number that designates the overall scale of compression.

Effects on Barcode Recognition

The recognition of barcode symbols relies on the analysis of small image details, and requires the detection of edges, and the measurement of edge to edge distances or of feature dimensions. Loss of edge information leads to lower symbol readability, over and above any issues of resolution and focus. DataMatrix is the very sensitive symbology to these effects because it requires decoding in both axes. PDF417 and 1D barcodes primarily decode along a single axis

Though JPEG compression effectively destroys information, ClearImage barcode recognition engines employ proprietary algorithms to recover the barcode from damaged symbols, even those that were reconstructed from highly compressed JPEG files. Nevertheless, the speed of recognition is reduced, as the software has to employ more elaborate computations to recover the image and data. Furthermore, at any given resolution and quality, there will be some images that are too far gone, and will not be recoverable.

2D barcodes (DataMatrix and PDF417) include Error Correction Codes (ECC) making them inherently more reliable. The ECC can enable the recovery of data from damaged symbols, after the symbol is reconstructed. The effectiveness of the ECC depends on the amount of ECC octets included in the symbol. When creating a PDF417 symbol, the user can select the ECC Level. In the case of DataMatrix symbols, the ECC level is inherent in the symbol. However, selection of larger matrix sizes offers more protection than smaller symbols, even as it wastes data space.

Recommendations

To the extent that your process allows you to do so, we suggest a number of steps that can be taken to improve recognition speed and rate:

- If the color or grayscale image bitmaps from the scanning driver are available in memory, use them rather than the compressed files.
- If the color or grayscale images must be read from file, use uncompressed or low loss compression (BMP, Uncompressed TIFF or JPEG with near 100% Quality factor) formats to create the file. You can set the formats and Jpeg Quality Factor in the Options or Preferences menus of the Save File dialogs in most image applications. Digital cameras usually offer a means to control the "Quality" parameter in their setup mode.
- If you are forced to use highly compressed JPEG files then:

    - Increase the module size for printing the symbol.
    - Increase scanning resolution.
    - Use 2D barcodes to take advantage of Error Correction Codes.
    - Use higher level of ECC to assure sufficient amount of correction data.

Compression and Resolution

The following examples visually demonstrate the effects of scanning resolution and JPEG compression. They were obtained by scanning and compressing the following DataMatrix symbol, shown here in it's actual size ➔

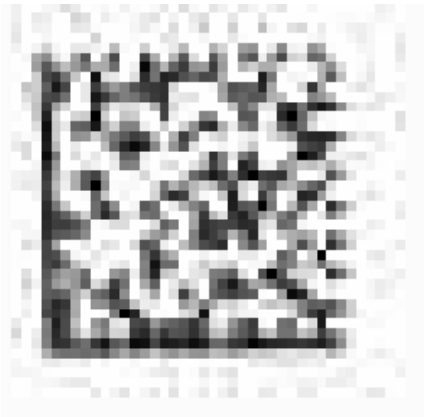
The Images were sufficiently magnified to make the scanning and compression artifacts visible to naked eye.

Observation made form these small set can be summarized this way:

- Scanned at 100 dpi, the symbol is at the very low end of readability.
- Additional JPEG compression severely hampers symbol readability
- At 150 dpi readability is significantly improved vs. 100 dpi image.
- 150 dpi scan at higher degree of JPEG compression (represented by low Quality Factor) is more readable then uncompressed 100dpi image

Scan Examples

100 dpi uncompressed image. Due to low resolution image has number of gray pixels, rather than well defined barcode elements.



100 dpi scan compressed at 50 quality. Image became fuzzy with extensive compression blocking artifacts, confounding the boundary of barcode elements.

150 dpi uncompressed. Higher resolution results in much sharper definition of barcode elements.

150dpi scan compressed at 0 quality. Image is fuzzy, and many artifacts are visible. However it is more readable than 100dpi uncompressed image.