RC 22014 (98866) 28 March 2001 **Engineering & Technology**

Research Report

Enhanced JPEG Compression of Documents

Ravi Prakash, Joan L. Mitchell[1], David A. Stepneski
IBM Payment Solutions
8501 IBM Drive
Charlotte, NC 28262

[1]IBM Research Division
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Enhanced JPEG Compression of Documents

Ravi Prakash[1], IEEE Member, Joan L. Mitchell[2], IEEE Fellow, and David A. Stepneski[1]

[1] IBM Payment Solutions, 8501 IBM Drive, Charlotte, NC USA 28262
[2] IBM T.J. Watson Research Center, Yorktown Heights, NY USA 10598
on loan to IBM Printing Systems Division, Boulder, CO USA 80301
E-mail: rprakash, joanm, dastep @us.ibm.com

ABSTRACT

The JPEG standard was designed for compression of photographic digital images, but it also works well on digitized documents with only a limited number of shades of gray. For documents in which compression and legibility are more important than preserving all of the intermediate values, preprocessing the images to reduce their dynamic range can enhance JPEG compression, as it selectively discards some noise. If modified quantization tables are substituted for the encoding quantization tables in the JPEG compressed data stream, an unchanged JPEG decoder can restore the dynamic range and increase image contrast. Graphs of the compressed size in bytes vs. dynamic range after the application of different dynamic-range-reduction techniques are given for both Huffman coding and arithmetic coding. Examples of the reconstructed front and back sides of a check with normal processing and enhanced compression are shown.

1. INTRODUCTION

With the Internet becoming an integral part of life, the ability to cost-effectively store and transmit photographs is critical. Today, one can hardly find a web page that does not contain some Joint Photographic Experts Group (JPEG) <1, 2> compressed photographs and illustrations. Since these compressed images need to travel on the information superhighway and are increasingly being contained in e-mails, document data size has surfaced as a key factor in ease of use.

The US Banking Industry is attempting to make available on-line to their customers, both commercial and individuals, images of their financial documents such as checks (cheques). In the banking world, a typical digital check record consists of a header followed by compressed image segments of the front and back of the check, an average total record size of 50 kilobytes. There are approximately 80 billion checks written per year in the United States. Saving a copy of these images for the required seven years translates into 28,000 trillion bytes of compressed data saved on write-once archival media. Not only is the total storage requirement vast, retrieving and transmitting these records as needed can also place a tremendous burden on the server and network. Thus, even a modest 10% increase in the compression of documents would offer great benefit to the financial industry. Assuming that the on-line storage available is not changed, an improvement in compression extends the time during which images are rapidly accessible.

Document legibility is also important. Since the JPEG baseline (DCT-based) compression is a lossy process, increased compression may remove vital information. However, documents in which some of the intermediate tones are not important to the legibility of critical information lend themselves to selective dynamic range reduction in a preprocessing step before JPEG encoding. If the quantization tables are replaced after encoding with scaled-up quantization values, any JPEG decoder can restore the dynamic range.

In this paper, the authors describe some experiments in reducing the dynamic range of check images and present some preliminary results. Section 2 illustrates some of the types of dynamic range reduction. Section 3 shows the effect of each type of range reduction on compression for both Huffman and arithmetic entropy coding. Section 4 describes how to restore the dynamic range. Section 5 shows some reconstructed images.

2. RANGE REDUCTION

Figures 1a, 1b, and 1c illustrate types of mappings which may be used to reduce the dynamic range by clipping, range scaling, and coring. The input and output pels are 8-bit values with 0 representing black and 255 representing white. In all cases, the output range is reduced compared to the input range. In real applications all three methods may be mixed to produce the best output quality for the range-reduced image.

Image clipping can be done from the white side, the black side, or both sides of the image. For example, clipping the darkest values to 16 means all pels having an intensity value below 16 are given a value of 16. Similarly, clipping the brightest values to 240 changes all pels with intensity values greater than 240 to 240. Figure 1a shows clipping the lowest 16 black values and the highest 32 white values.

Range scaling is defined as a linear reduction of output values relative to the input values. Figure 1b shows a 2:1 linear scaling of the original 256 values down to a range of only 128 values. Note that the midpoint has remained fixed so the output values range from 64 to 191.

Image coring is illustrated in Figure 1c. The center 32 values near 128 have been replaced with a constant value of 128. Then the values on both sides are shifted toward 128 to preserve continuity in the histogram. Values from 0 to 112 are shifted up to 16 to 128. Similarly, all values from 144 to 255 are shifted down to 128 to 239. If the coring is centered on a peak in the image histogram, significant noise is removed from the image when the variations in the peak are collapsed to a single value.

3. EFFECT ON COMPRESSION

The financial industry has standardized on a minimum resolution for a document's gray scale images (such as the front and back of a check) of 100 dots per inch (dpi) or 4 dots per mm. Figure 2 shows gray scale images of the front and back of a personal check processed to 120 dpi, 20% above this minimum resolution. The back side of the check has the endorsement field blanked out. All compression numbers were collected on the original information without this modification.

Figure 3 gives the histograms for these original images. Note that the background peak on the front side of the check is much broader than that on the back side, and that the significant information on the front of the check is generally higher contrast than the endorsements on the back.

Preliminary experiments showed that reduction in image dynamic range has a favorable impact on the compressed size. Effects of clipping, range scaling and coring on the compressed image sizes of these images are shown for both Huffman coding and arithmetic coding, for the front side of the check in Figures 4a and 4b and for the back side of the check in Figures 5a and 5b. The drop in compressed image size from extreme clipping indicates that most of the image content has been removed. However, these curves can be used to optimize mixtures of the range reduction techniques to preserve legibility.

The Huffman coding uses the JPEG baseline compression algorithm with the example luminance quantization table (Table K.1) and fixed Huffman tables (Tables K.3 and K.5) found in the standard <1, 2>. The DC values at the far left of the graphs (where all pels have been mapped to a single value) will be a constant. The Huffman code for a constant DC value is two bits long. The End-of-Block (EOB) will occur immediately since there will be no AC coefficients in a flat image. The EOB code is four bits long. The average of 6 bits/block times the 3600 blocks in the check front and the 3690 blocks in the check back yields a lower limit for the fixed Huffman coding of 2700 bytes and 2767 bytes respectively. The header information to send the quantization table (68 bytes) and the Huffman tables (204 bytes) adds another 272 bytes. In the real system this information is not stored with every image.

The arithmetic and Huffman coding curves show similar relative shapes, except that the arithmetic coding limit on the left edge approaches zero. The header is smaller since the arithmetic coding tables are conveyed in just a few bytes compared to a few hundred for the Huffman tables. Scaling the dynamic range had an approximately linear effect on the compression. The histograms of the original raw images can be used to estimate from the cumlative number of pels affected when clipping will start to affect the number of compressed bytes.

4. RANGE RESTORATION

Figure 6 shows a flow chart of the encoder's processing for enhanced JPEG compression. The first block is an optional histogram. A histogram allows custom image parameters to be determined, e.g., the background peak and maximum range of the image. The second block is an optional preprocessing step which could use smoothing filters or other noise removal techniques. Using both of these steps allows custom clipping and coring <3> before range scaling.

The next block shows determining the desired new range. This may be done per image based on the optional histogram and/or preprocessing, or it may be preset for a large number of images. An example of presetting the range reduction is to pick a power of two such as a reduction of 2:1.

Once the new range and the encoder's quantization table (Q-table1) are known, the decoder's quantization table (Q-table2) can be approximately calculated by scaling Q-table1 by the desired range expansion. The desired range expansion may be more than the original range reduction. For example, if the histogram shows that the image started with poor contrast (i.e., less than the maximum range), the desired output range can correct for this by scaling up by more than the range reduction. In general, people tend to prefer higher contrast, and so some contrast enhancement can be built into Q-table2.

The JPEG standard specifies that the quantization values are restricted to 1 byte each for images with 8 bits per sample per component. For large reductions in the dynamic range simply scaling the Q-table1 quantization values is likely to exceed the maximum 1-byte value of 255 for many of the quantization values. Some consideration can be given as to whether to simply clamp such values in Q-table2 to the maximum value of 255 or whether to start with the 255 value and determine the Q-table1 values by dividing 255 by the reduction factor.

The next block reduces the dynamic range of the image. For text documents, we are likely to want to reduce the dynamic range rather severely since the originals are often high contrast black/white images and we are not particularly interested in preserving the intermediate values. For checks, the safety pattern and background pictures may still be somewhat important.

It is important to keep the final reduced range approximately centered within the desired final, reconstructed range. A simple shift right to reduce the range moves the midpoint of the range; upon decoding, only one side of the AC coefficients has the freedom to re-expand the dynamic range fully and the image is much too dark.

The next blocks show JPEG compression of the modified-range image using a first quantization table (Q-table1). However, the quantization table stored within the JPEG Define Quantization Table (DQT) marker is the modified Q-table2. Any JPEG browser will be able to restore the range using the modified Q-table2. The final block in Figure 6 indicates that the JPEG encoded modified image is either transmitted or stored.

5. SOME RESULTS AND DISCUSSION

Figures 7a and 8a show reconstructed images encoded and decoded with the same quantization table (Q-table1). Figures 7b and 8b show the same images with dynamic range reduction in the encoder, encoded with Q-table1, and decoded with a different quantization table (Q-table2). Q-table1 and Q-table2 were selected to give approximately the same quality. While the two images are similar in appearance, the second pair of images have more than 20% fewer compressed bytes both when compressed with Huffman coding and arithmetic coding. The enhancement in compression is slightly more for the arithmetic coding, which can take better advantage of the reduction in noise.

Reduction in image dynamic range of documents delivers increased compression. Finally, the dynamic range is recovered by adjusting the decoder's quantization table to produce an image similar in appearance to the original. For a given level of desired compression, the various parameters associated with dynamic range reduction can be optimized to achieve the highest possible level of image legibility, using the altered quantization table.

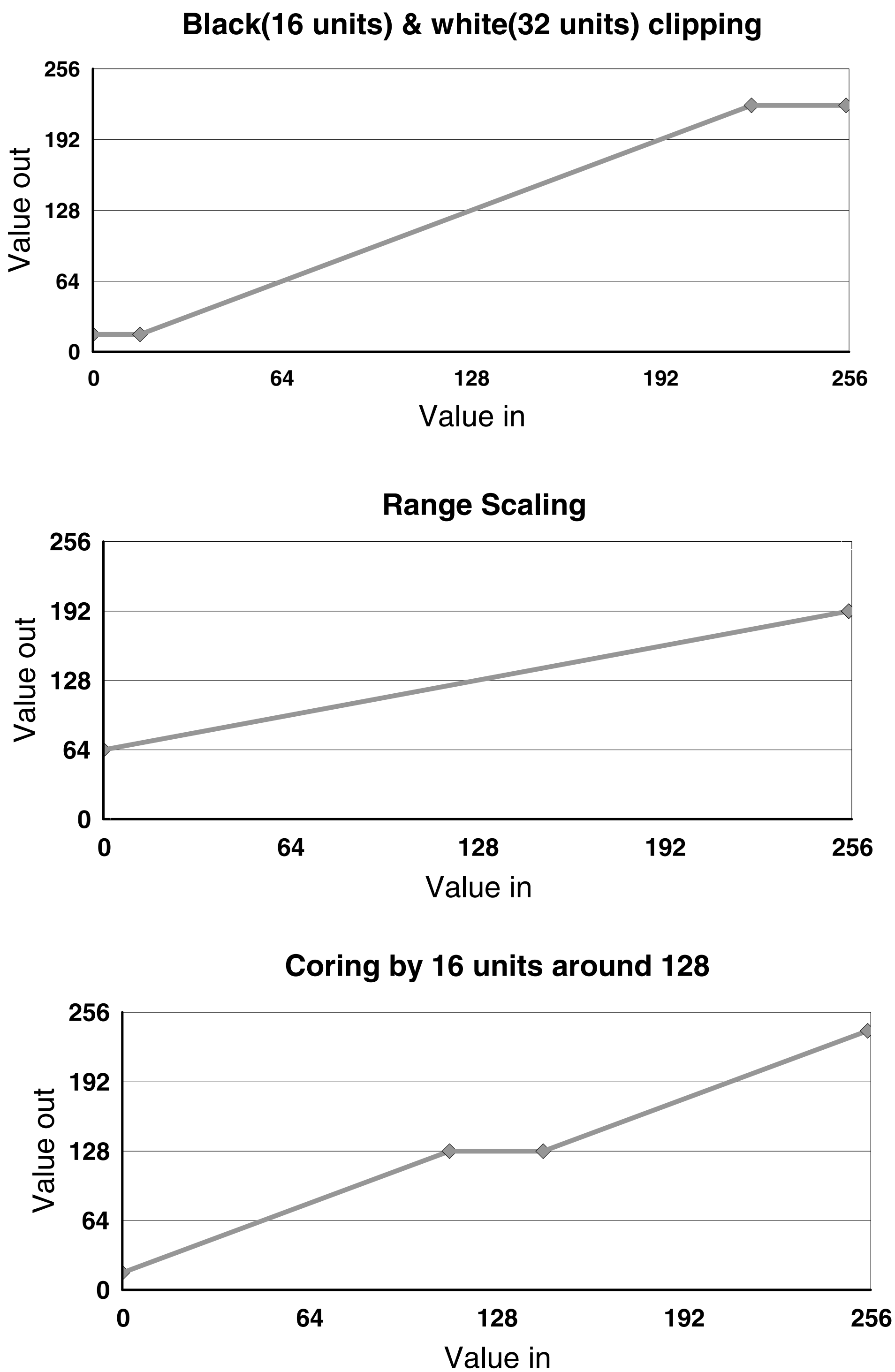
The JPEG standard specifies that any reconstructed values outside the allowed range are expected to be clamped at the appropriate boundary. Since human observers tend to prefer high contrast images of documents, this internal clamping can be used to create a higher contrast image if Q-table2 is designed to overcompensate for the original range reduction. In the process, some edge quantization noise may be clamped away.

In the banking industry, the digitized, compressed documents may be immediately stored in write-once media for legal reasons. Often an additional copy is kept on-line for several months to enable rapid Internet access. To allow more images to be stored online for a longer period of time, the highest quality compressed images could be kept for a shorter period such as a month. Then the images could be decoded, range reduced (perhaps for the first time), and re-encoded. Further decompression, additional dynamic range reduction, and recompression cycles would allow images to be kept for even longer periods of time. For most requests the degraded image is still legible enough to rapidly confirm the contents of the image. The legal version could still be accessed from the write-once media if necessary.

6. REFERENCES

1. ITU-T Recommendation T.81 | ISO 10918-1: Digital Compression and Coding of Continuous-Tone Still Images, Part 1.
2. W.B. Pennebaker and J.L. Mitchell, *JPEG: Still Image Data Compression Standard,* Van Nostrand Reinhold, New York, 1993.
3. European Patent Application EP 188193, W.B. Pennebaker and J.L. Mitchell, "Method and Apparatus for Processing Image Data," published 7/23/86.

Figure 1. Range reduction. a. Clipping 16 black values and 32 white values. b. Range scaling by a factor of 2. c. Coring 32 values centered at 128.



Enhanced JPEG Compression of Documents

Figure 2. Grayscale images. a. Check front (720 pels X 320 lines = 230.4 kbytes).
b. Check back (720 pels X 328 lines = 236.16 kbytes).

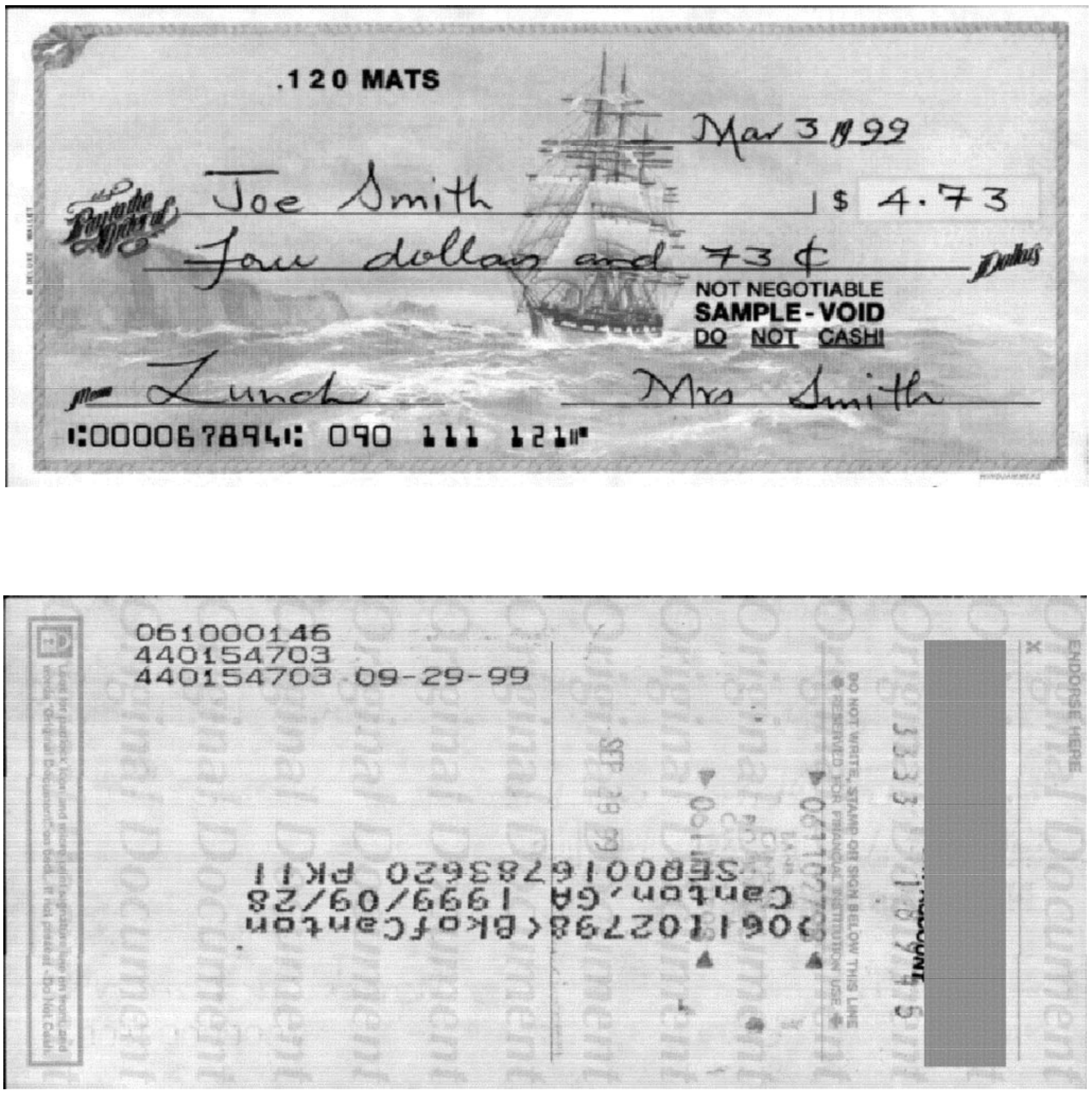
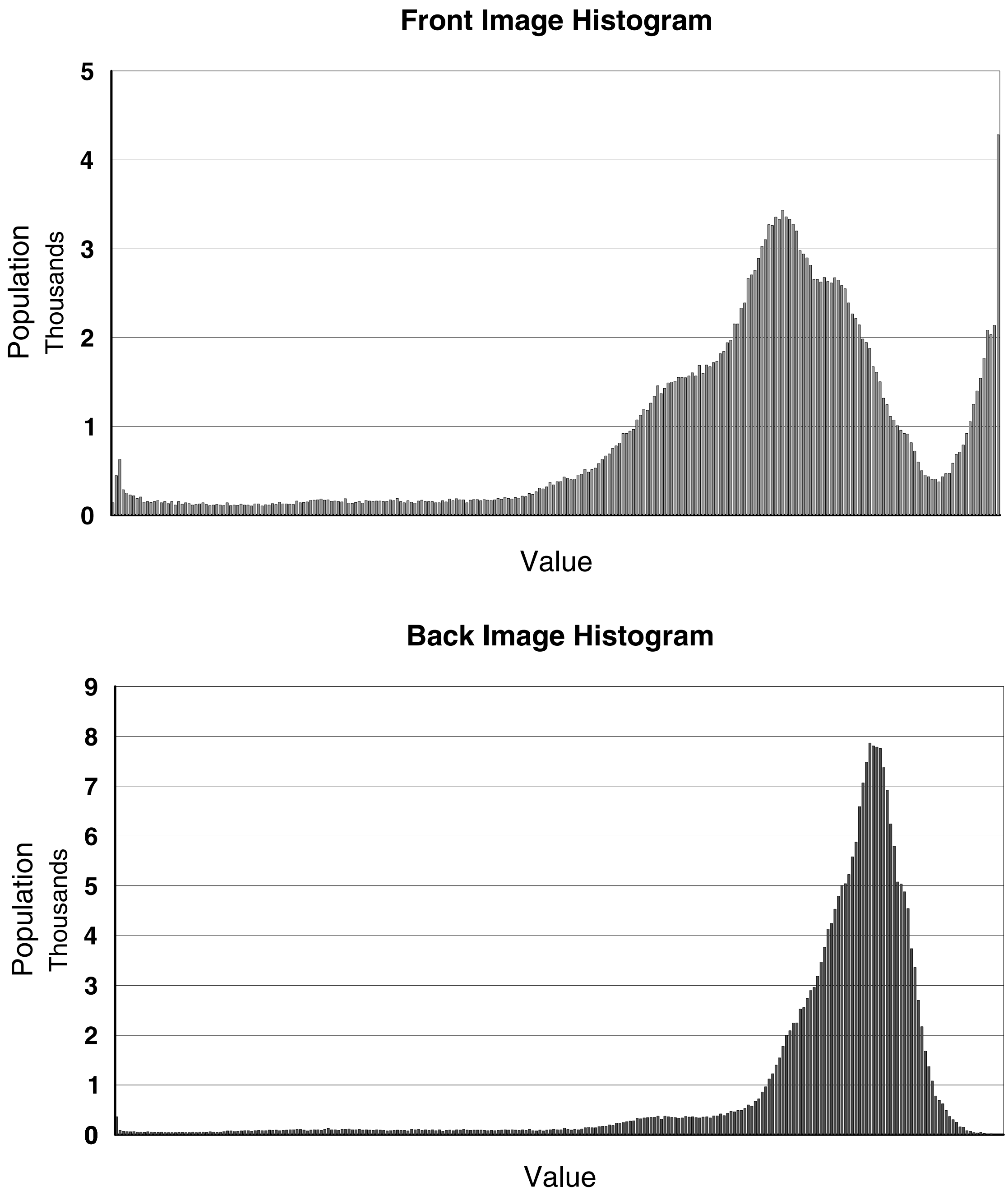
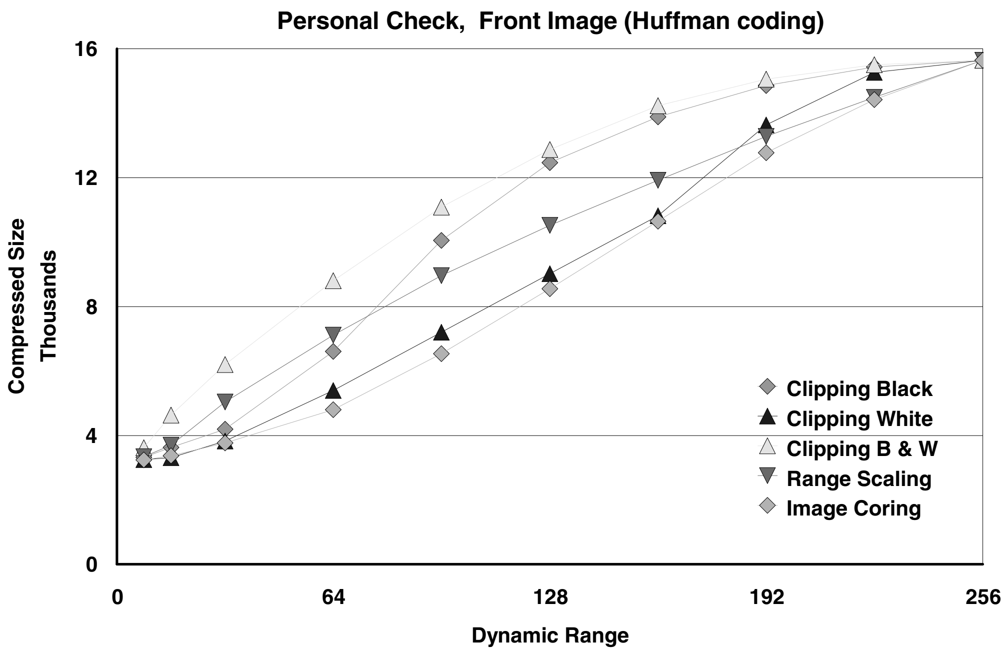


Figure 3. Histogram of original data. a. Check front. b. Check back.



Enhanced JPEG Compression of Documents

Figure 4. JPEG compressed size (bytes) versus dynamic range, front image. a. Huffman coding. b. Arithmetic coding.



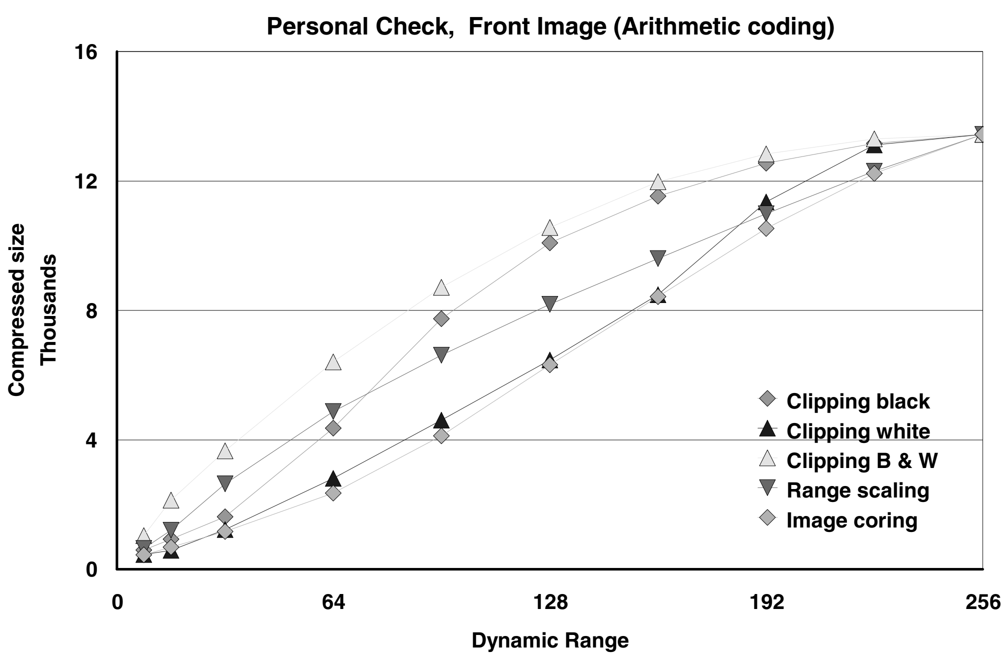
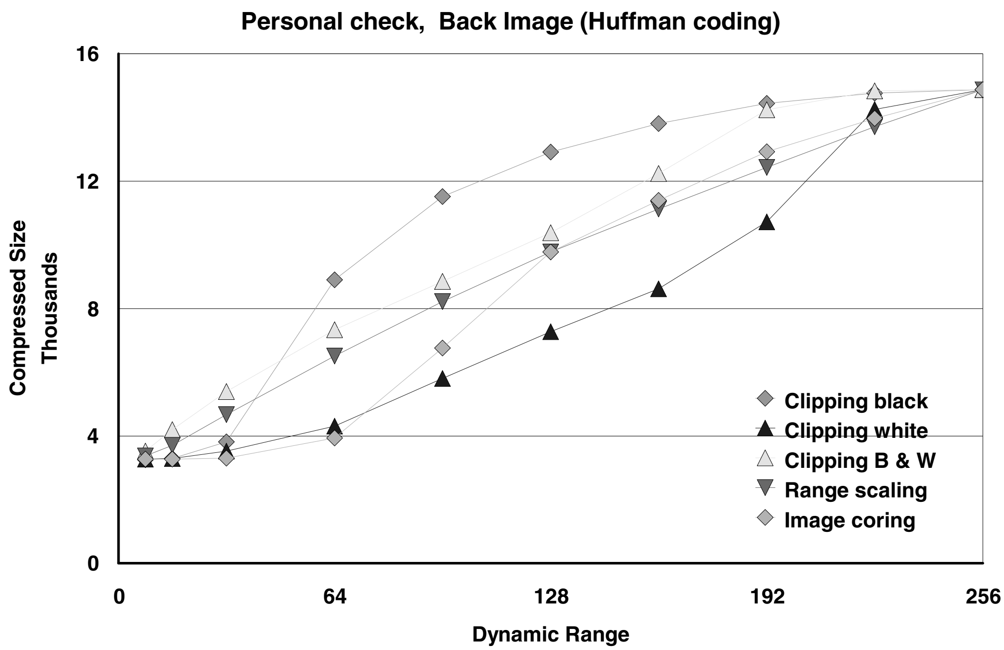
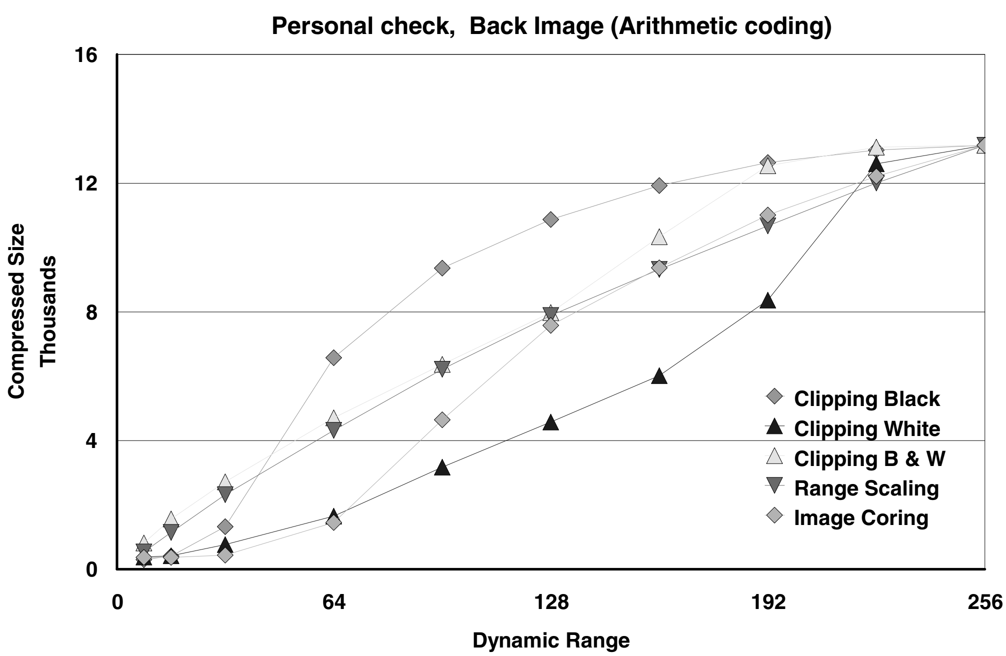


Figure 5. JPEG compressed size (bytes) versus dynamic range, back image. a. Huffman coding. b. Arithmetic coding.





Enhanced JPEG Compression of Documents

Figure 6. Flow chart of enhanced JPEG compression of documents.

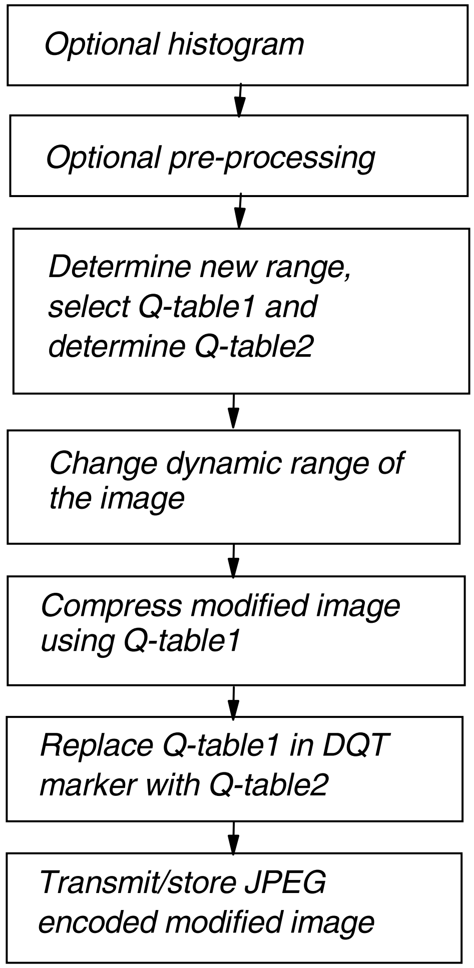


Figure 7. Reconstructed front images: a. After normal processing using Huffman coding (15,650 bytes) or arithmetic coding (13,445 bytes). b. After enhanced compression, using Huffman coding (12,805 bytes) or arithmetic coding (10,470 bytes).

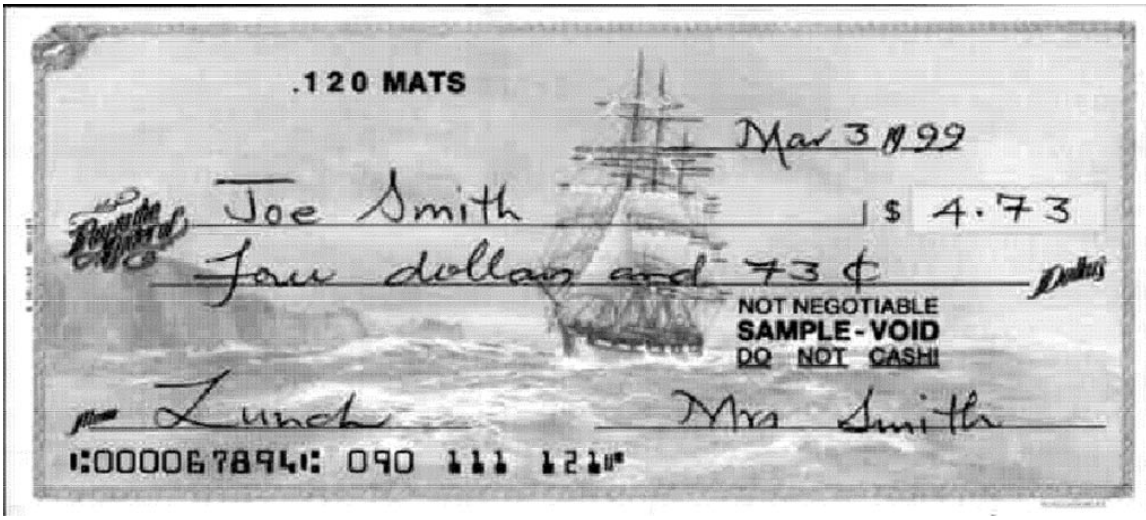


Figure 7a.

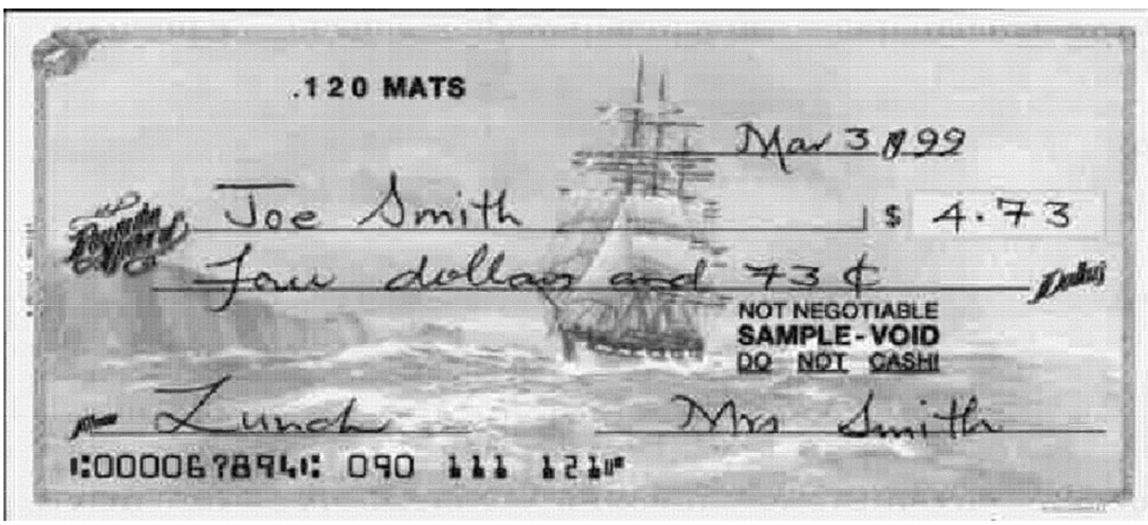


Figure 7b.

Enhanced JPEG Compression of Documents

Figure 8. Reconstructed back images: a. After normal processing using Huffman coding (14,832 bytes) or arithmetic coding (13,179 bytes). b. After enhanced compression, using Huffman coding (12,044 bytes) or arithmetic coding (10,266 bytes).

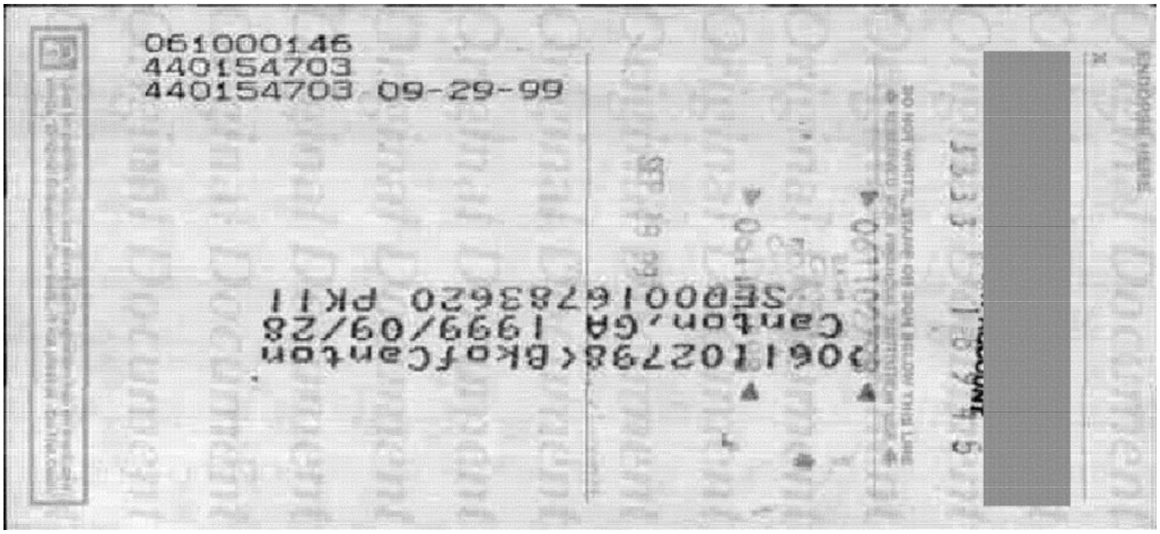


Figure 8a.

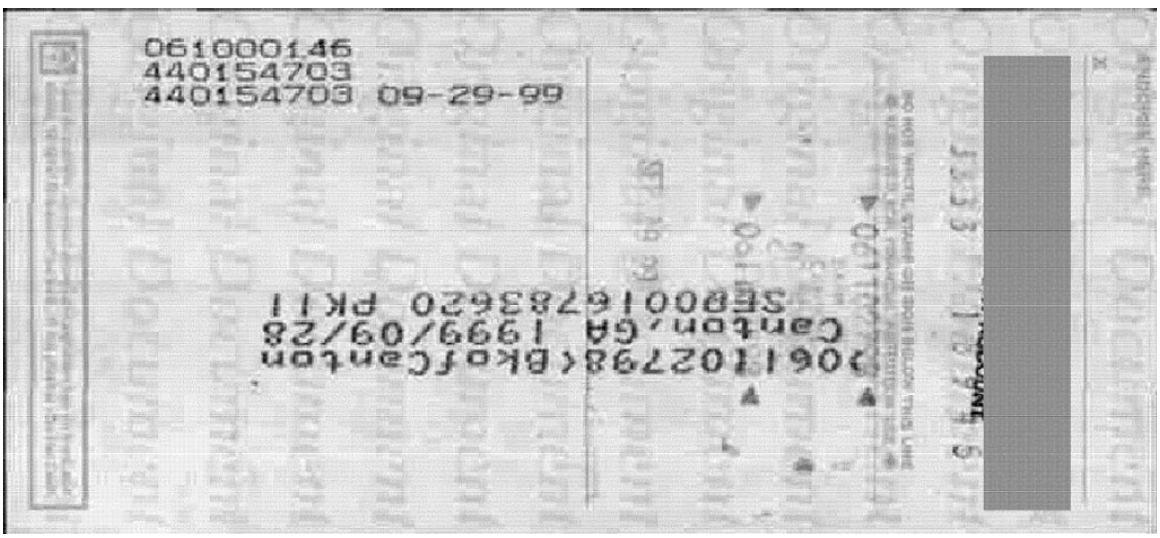


Figure 8b.