EFFICIENT AND RELIABLE DYNAMIC QUALITY CONTROL FOR COMPRESSION OF COMPOUND DOCUMENT IMAGES

Amir Said
Hewlett-Packard Laboratories, Palo Alto, CA, USA (said@ieee.org)

ABSTRACT

Compound images contain a mixture of natural images, text, and graphics. They need special care in the use of compression because text and graphics cannot withstand the significant distortion that is acceptable for natural images. One solution is to identify the areas that need to have higher settings in a dynamic quality control scheme, as supported by JPEG-SPIFF and JPEG 2000. We propose a scheme to identify those regions using a discrimination function that works with a non-linear transform to reliably identify edges, and at the same time avoid false positive detection on regions with complex patterns. It does so by exploiting the properties of histograms of coefficients of this block transform, and their entropy function, which we show can be computed efficiently via table look-up. Experimental results demonstrate the performance of the new scheme, compared to other methods in the literature.

1. INTRODUCTION

Compound images are those like the example shown in Fig. 1, which contains a combination of natural images (photos), text, and graphics. This type of image occurs in many important applications, like document imaging and printing [2-8] and in remote visualization [10].

The standard transform-based (e.g., discrete cosine or wavelet) image compression methods effectively exploit the properties of the natural images and of the human visual system, and are able to change natural images considerably, while keeping an excellent visual quality.

The images containing text and graphics, on the other hand, have edges that are harder to represent with the linear transforms, and consequently many noticeable artifacts tend to appear around edges. What motivates the development of special compression for compound documents is the severity of this problem in the text regions. The complex structure of the many text edges produces an accumulation of artifacts around text, and the problem is made worse by the fact that the visual quality expectations for text are very different from those for natural images. People expect text to be crisp, and we easily perceive how artifacts reduce legibility and reading speed, and lower the overall document quality.

One solution is to use classification with dynamic quality, i.e., identify the areas that may contain text and graphics, and use higher quality settings during their compression. In fact, for high-quality printing applications it is best to use lossless compression in those regions [4]. Other applications have less stringent requirements, and it may be best to use a single coding standard [2,5,6]. Both the JPEG-SPIFF [1] and JPEG 2000 [9] image coding standards support some form of dynamic quality control.

In this paper we analyze what are the desirable properties of method for classifying the image regions. Based on this analysis we propose a discrimination function that uses a non-linear transform for edge detection, and exploits the properties of the entropy of the coefficient histogram. We show that one of the properties enables a very low complexity implementation of this function, using only table look-up for the computations.

Extensive experimentation shows the accuracy of this technique, identifying edges in text regions, but ignoring edges in complex patterns, where artifacts are less visible.

Fig. 1 – Example of a compound image.

2. CLASSIFICATION TECHNIQUES

There are sophisticated techniques that can be used for document analysis and understanding (see, for instance, [7,8] and their references). However, we are considering a quite different problem, because we are interested only in the compression consequences. For instance, for our purposes we do not consider ignoring blurred text as an incorrect classification if it will not produce objectionable visual artifacts. Furthermore, we look for a method that is simple, reliable, and has requirements that are comparable and compatible with the compression method [5,6].

The classification can be based on data computed in the spatial domain, or on the results of some transform [2,6,7]. It is important to note that the latter does not mean lower complexity. For instance, in hardware the analysis can be done in parallel with the transform [5].

Since one of the main characteristics of text and graphics is the presence of edges, methods that use transforms that do not have good edge discrimination are prone to classify noisy and patterned regions as text. Thus, they increase the quality settings—and thus use significantly more bits—in exactly the regions that can be heavily compressed because the artifacts are hard to see.

Memon and Tretter [5] also use the results of a transform for classification, but one that is more suited for finding edges, instead of coding. In this paper we show that we can improve the reliability of their approach by efficiently analyzing the histogram of coefficient magnitudes, through the computation of its entropy.

3. ENTROPY OF BLOCK DIFFERENCES

As explained above, we need a transform to detect blocks that have edges and few luminance levels, and reject blocks with edges plus complex textures and noise. Let us represent the luminance of each pixel as $p_{i,j}$. We can get information about the block activity from the horizontal and vertical absolute differences

$$d^{h}_{i,j} = \left| p_{i,j} - p_{i,j-1} \right|, \quad d^{v}_{i,j} = \left| p_{i,j} - p_{i-1,j} \right|, \qquad (1)$$

which define our non-linear transform.

In [5] the classification is based on the effective sequential analysis of the sum of the absolute differences, the maximum absolute difference, and the number of zero differences. However, by aggregating all the absolute differences in a sum, useful information is completely lost. For instance, we cannot reliably identify if we have a few large or many small differences.

Let h_n, $n = 0, \ldots, S-1$, be a set of non-negative numbers corresponding to a histogram of absolute differences (both vertical and horizontal) in an $N \times N$ image block, and define the number of computed differences as

$$T = \sum\nolimits_n h_n = 2N(N-1)\,. \qquad (2)$$

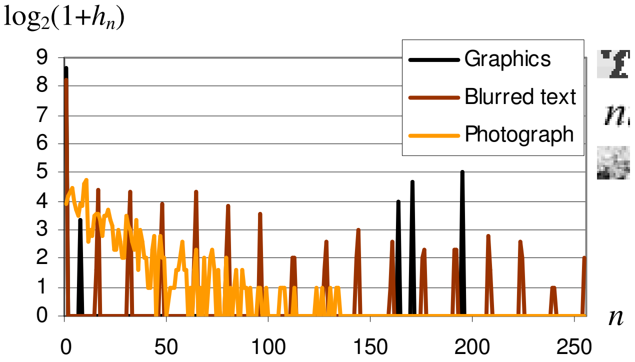


Fig. 2 – Typical histograms of difference magnitudes.

The histogram provides much more detailed information, which can improve the reliability of the block classification. Fig. 2 shows the type of distribution normally expected for some types of block content (note the logarithmic vertical scale). Blocks that are part of photos or complex graphic textures tend to have histograms that are peaked at zero with a somewhat slow decay. Blocks with text and graphics normally have histograms containing just a few peaks, while blurred text has many peaks [8].

We want to use this property of the histograms to identify the text blocks, but we also want to minimize the computational effort analyzing the histogram. One function that can do this classification is the entropy

$$\psi(\mathbf{h}) = \sum_{h_n \neq 0} \frac{h_n}{T} \log_2\left(\frac{T}{h_n} \right), \qquad (3)$$

which is small for histograms with peaks (text, graphics, smooth areas), and large for relatively flat histograms (complex patterns). For example, its value is 0.96, 2.5, and 5.9 for, respectively, the graphics, blurred text, and photograph histograms of Fig. 2. The entropy also has the advantage of being much more dependent on the shape of the histogram then the actual values or the edge shapes. For instance, the graphics corresponding to the histogram in Fig. 2 has quite complex edges, but its histogram entropy is comparable to a block with a single straight edge. The same applies to blocks with edges covering many or just a few pixels.

4. EFFICIENT COMPUTATION

Since the entropy function is defined using logarithms, divisions and multiplications, it appears to be computationally *very* costly. Instead, it is quite the opposite, because we can use the fact that (3) is equivalent to

$$\psi(\mathbf{h}) = \log_2(T) - \frac{1}{T} \sum_{h_n \neq 0} h_n \log_2(h_n)\,, \qquad (4)$$

plus the fact that in an $N \times N$ image block we have $0 \le h_n \le T$.

Consequently, we can pre-compute all the possible terms using fixed-point representation. If we use 10 bits for the fractional part, we first create a table of integers

$$F(n) = \begin{cases} 0, & n = 0, 1, \\ \left\lfloor \dfrac{1024\, n \log_2(n)}{T} + \dfrac{1}{2} \right\rfloor, & n = 2, 3, \ldots, T, \end{cases} \qquad (5)$$

and then compute (4) using only table look-up and integer arithmetic as

$$\psi(\mathbf{h}) \approx \frac{1}{1024}\left[F(T) - \sum_n F(h_n) \right]. \qquad (6)$$

The computation of (6) can be done while clearing the histogram values for the next block.

5. SCALED ENTROPY

The entropy function provides a good classification of the histogram shape. However, we do not want to positively identify the smooth blocks, which also have a low-entropy histogram. Fortunately, smooth blocks are easily identified because all h_n are small. We can use histogram to compute a more reliable estimate as

$$\mu(\mathbf{h}) = \max\{k : \sum_{n \ge k} h_n \ge t_{\text{top}}\}. \qquad (7)$$

For example, if we use $t_{\text{top}} = 1$, we just have the maximum difference magnitude. In our tests we used $t_{\text{top}} = 4$. We can incorporate this factor in our discrimination function by defining scaled entropy as

$$\Phi(\mathbf{h}) = \frac{a[\psi(\mathbf{h}) + b]}{\mu(\mathbf{h}) + c}, \qquad (8)$$

where a, b, and c are constants that normalize the entropy, and change the sensitivity of the classification.

6. COMPARATIVE EVALUATION

In [5] we can find how a discrimination function can be translated into quality settings for the JPEG part 3 coding standard [1]. The full implementation of the classification normally uses several different values, in a series of tests, before reaching a decision. These details are beyond the scope of this summary. Nevertheless, we are able to show experimental results confirming the expected superior performance of the histogram analysis.

Fig. 3 shows the application of different discrimination functions to 8×8 blocks in a series of four monochrome images, shown in the first column. The next columns show a mapping of the function results to a gray level in each block. For consistency we rescaled the functions to obtain maximum contrast, and in such a way that brighter blocks represent those that need higher quality settings.

The second column of images shows the function computed via logarithms of the 8×8 DCT coefficients [6], and the third column a sum of the difference magnitudes [5]. The fourth column contains the result of the proposed scaled entropy function. We can clearly see that our analysis of the histogram yields a much more reliable identification of the regions with text and graphics. We can see it is less sensitive to noise and complex image patterns, as it can also flag the photo of text, and in the clock image.

7. CONCLUSIONS

We propose a new technique for identification of text and graphics in compound document images, which extends the work of Memon and Tretter [5]. Our classification is based on a histogram of the magnitudes of vertical and horizontal pixel differences. We analyze how the histogram shape is different for graphics and text and natural images, and propose the use of a scaled entropy function to enable the reliable identification. In addition, we show that the computation of this function can be done very efficiently using only table look-up and integer arithmetic. Simple experimental results provide a visual indication of the improved performance of the method.

REFERENCES

[1] ITU-T Recommendation T.84, *Digital Compression and Coding of Continuous-Tone Still Images: Extensions*, International Communication Union, Geneva, July 1996.

[2] N. Chaddha and A. Gupta, “Text segmentation using linear transforms,” *Proc. Asilomar Conf. Signals, Systems, Computers,* pp. 1447–1451, 1996.

[3] R.L. de Queiroz, “Compression of compound documents,” *IEEE Int. Conf. Image Processing*, pp. 209–213, Oct. 1999.

[4] A. Said and A.I. Drukarev, “Simplified segmentation for compound image compression,” *IEEE Int. Conf. Image Processing*, pp. 24–28, Oct. 1999.

[5] N. Memon and D. Tretter, “Method for variable quantization in JPEG for improved perceptual quality,” *SPIE Image and Video Communications and Processing Conf.,* Proc. SPIE vol. 3974, pp. 24–34, April 2000.

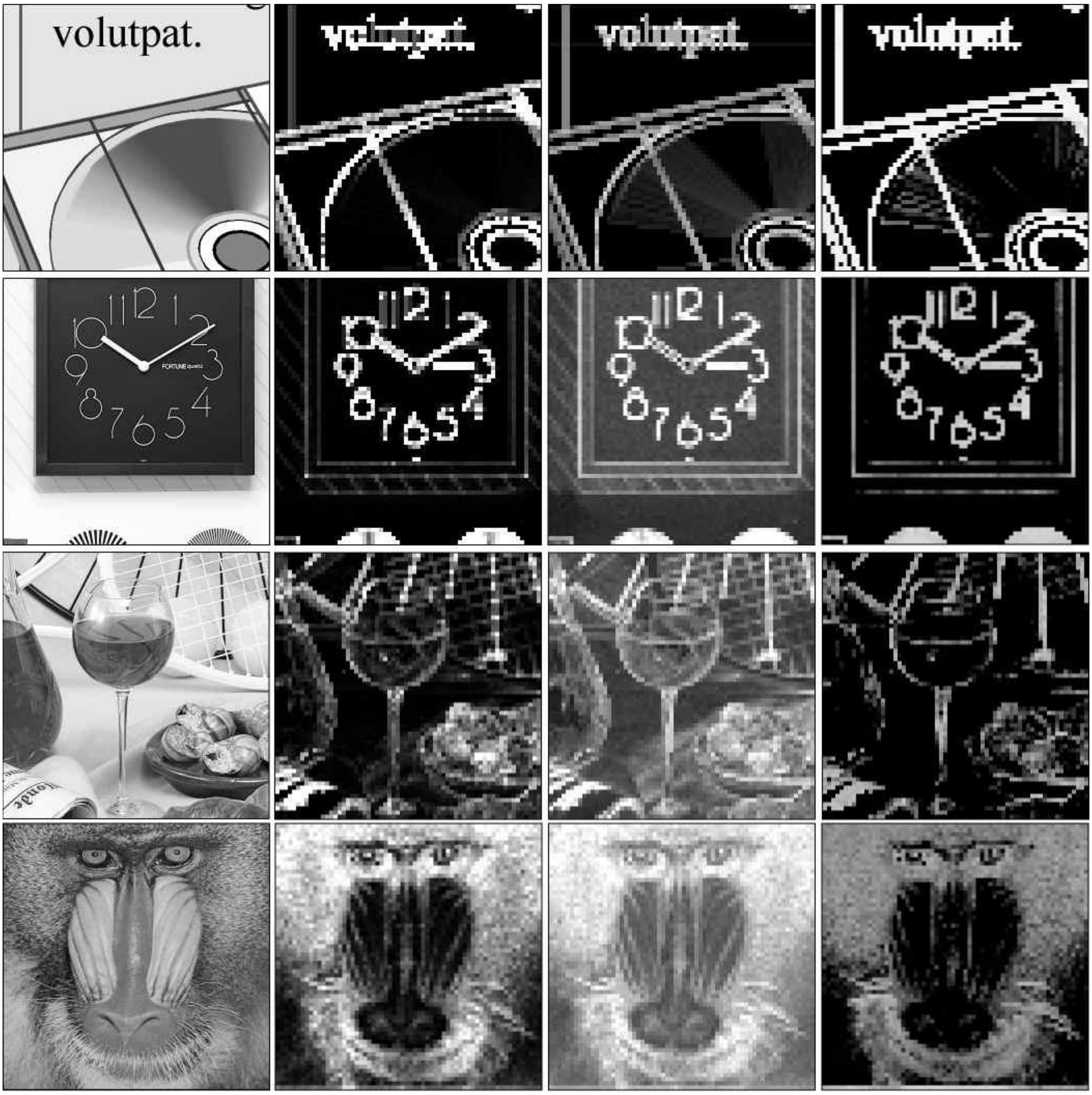
[6] K. Konstantinides and D. Tretter, “A JPEG variable quantization method for compound documents,” *IEEE Trans. Image Processing,* pp. 1282–1287, July 2000.

[7] J. Li and R.M. Gray, “Context-based multiscale classification of document images using wavelet coefficient distributions,” *IEEE Trans. Image Processing,* vol. 9, pp. 1604–1616, Sept. 2000.

[8] H. Cheng and C.A. Bouman, “Multiscale Bayesian segmentation using a trainable context model,” *IEEE Trans. Image Processing,* vol. 10, pp. 511–525, April 2001.

[9] D.S. Taubman and M.W. Marcellin, *JPEG 200 Image Compression Fundamentals, Standards and Practice,* Kluwer Academic Publishers, Boston, 2002.

[10] A. Said, “Compression of compound images and video for enabling rich media in embedded systems,” *SPIE Visual Commun. Image Processing Conf.,* SPIE Proc. vol. 5308, pp. 69–82, Jan. 2004.



(a) (b) (c) (d)

Fig. 3 – Mapping of different classification functions to gray levels. The column (a) contains the original images, and the following columns contain the grayscale mapping of the function value in each 8×8 block. All functions scaled for maximum contrast, and to represent blocks that need higher quality settings as brighter. The images in column (b) were computed using the logarithm of the 8×8 DCT coefficients, in column (c) using a sum of the pixel difference magnitudes, and in column (d) using the proposed scaled entropy of the histogram of pixel-difference magnitudes.