OPTIMIZING BLOCK-THRESHOLD SEGMENTATION FOR MRC COMPRESSION

Ricardo L. de Queiroz[1], *Zhigang Fan*[2] *and Trac D. Tran*[3]

[1,2] Xerox Corporation 800 Phillips Rd., 128-27E, Webster, NY, 14580
[3] The Johns Hopkins University
[1] queiroz@wrc.xerox.com [2] zfan@crt.xerox.com [3] ttran@ece.jhu.edu

ABSTRACT

Compound document images contain graphic or textual content along with pictures. They are a very common form of documents, found in magazines, brochures, web-sites, etc. We focus our attention on the mixed raster content (MRC) multi-layer approach for compound image compression. We study block thresholding as a mean to segment an image for MRC. An attempt is made to optimize the block threshold in a rate-distortion sense. Rate-distortion curves are presented to demonstrate the performance of the proposed algorithm.

1. INTRODUCTION

Document compression is frequently linked to facsimile systems, in which large document bitmaps are compressed before transmission over telephone lines. There is now a focus on new standards to provide color facsimile services over the telephone network and the Internet [1]. See [2] for a short overview of document compression standards and techniques. When it comes to compound documents different compression algorithms may be applied to each of the regions of the document, by either segmenting the regions or by generating multiple image layers. The mixed raster content (MRC) imaging model [1],[3],[4], which is a multilayer model will be the focus of this paper.

MRC enables a multi-layer multi-resolution representation of a compound document, as illustrated in Fig. 1. The basic 3-layer MRC model represents a color image as two color-image layers (Foreground or FG and Background or BG) and a binary layer (Mask). The Mask layer describes how to reconstruct the final image from the FG/BG layers, i.e. to use the corresponding pixel from the FG or BG layers when the mask pixel is 1 or 0, respectively, in that position. Thus, the FG layer is essentially poured through the Mask plane onto the BG layer. Once the original single-resolution image is decomposed into layers, each layer can be processed and compressed using different algorithms. Layers may contain different dimensions and have offsets associated with them. The compression algorithm and resolution used for a given layer would be matched to the layer's content. The compressed layers are then packaged in a format, such as TIFF-FX [5] or as an ITU-T MRC [3] data stream for delivery to the decoder. At the decoder, each plane is retrieved, decompressed, processed (which might include scaling) and the final image is recomposed using the MRC imaging model.

Since multiple planes represent a single image, the amount of data to be encoded is multiplied. In each layer there is redundant data, i.e. each layer (FG or BG) may contain "unused" pixels, as the pixels in that position will be selected from the other layer. Those unused pixels can be replaced by any color in order to enhance compression, since they do not affect reconstruction. This is the function of the data-filling processors illustrated in Fig. 2 for the case of a 3-layer MRC approach. Note that, given the filling algorithm, from Fig. 2 one can see that the segmenter function is to find a binary mask for a given input, from which the data-filling processors can derive the output layers based on the input image. See [6] for details on data filling algorithms.

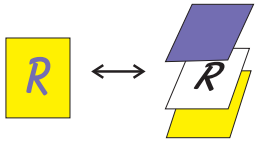

Figure 1. A 3-layer MRC imaging model.

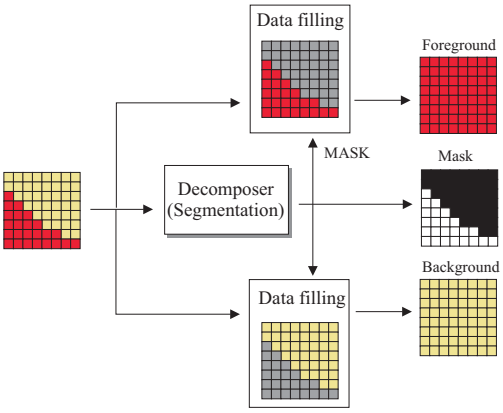

Figure 2. Diagram of a segmenter.

The image $x(i,j)$ is segmented generating the Mask $m(i,j)$. The data filling processors generate the image layers $L^{(BG)}(i,j)$ and $L^{(FG)}(i,j)$. The FG/BG planes are processed, compressed and combined into a single stream of bits which is transmitted to a receiver. The length of this stream is the rate R achieved by the coder. At the receiver, the streamed image data is then parsed and the FG/BG layers are decompressed into $\hat{L}^{(BG)}(i,j)$ and $\hat{L}^{(FG)}(i,j)$, while the Mask layer is supposed to be lossless encoded. The image is then recomposed as

$$\hat{x}(i,j) = m(i,j)\hat{L}^{(FG)}(i,j) + (1 - m(i,j))\hat{L}^{(BG)}(i,j)\ . \quad (1)$$

The distortion incurred by the compression is

$$D = \sum_{ij} \left(x(i,j) - \hat{x}(i,j)\right)^2 . \quad (2)$$

In the case of block-based segmentation the image is divided into blocks of 8×8 pixels and all quantities (e.g. x, L, m, D, R) are represented with subscript m to designate the m-th block. In this paper, the rate achieved for the block is not computed but estimated as

$$R_n = R_n^B + R_n^M + R_n^F, \tag{3}$$

where R_n^B, R_n^M and R_n^F are the estimated rates for compressing $L_n^{(BG)}(i,j)$, $m_n(i,j)$ and $L_n^{(FG)}(i,j)$, respectively. The reason for estimating as opposed to computing the actual rates is twofold. First, for the Mask block, it is very difficult (if not impossible) to determine the amount of bits a single block will generate since binary coders generally make use of run-lengths, differential positions, objects, etc. We simply estimate the mask rate by counting the number of horizontal transitions and apply a fixed average penalty (e.g. 7 bits per transition). So, for a block with N_t transitions,

$$R_n^M = N_t * \text{penalty}. \tag{4}$$

Second, even though R_n^B and R_n^F can be precisely computed for coders such as JPEG, it is conceivable that all the compressed data may be further subject to entropy coders. Here, the compressed layer files are collected using the Unix programs `tar` and `gzip`. Thus, the bit-stream is subject to Lempel-Ziv compression. Even though $D = \sum_n D_n$, the rate is only approximated, i.e. $R \approx \sum_n R_n$.

For a number of reasons better discussed in [2] we apply JPEG [7] to compress the FG/BG layers, without spatial scaling and using the same quantizer table $q(i,j) = Qq_d(i,j)$, where $q_d(i,j)$ is JPEG's default. The binary plane is compressed with MMR [8].

2. OPTIMIZED BLOCK THRESHOLDING

For a given image $\{x(i,j)\}$, using a particular coding quantizer Q_c, R and D will depend on $\{x(i,j)\}$, Q_c and $\{m(i,j)\}$. Here, we want to optimize the segmentation in an RD sense and, therefore, it is assumed a quantizer Q_d for the design phase, and an operating point λ to control the RD trade-off. Hence, $\{m(i,j)\}$ is a function of λ, Q_d and $\{x(i,j)\}$.

In block thresholding the mask is found as

$$m_n(i,j) = u(t_n - x_n(i,j) - 1) \tag{5}$$

where t_n is the block's threshold and $u(k)$ is the discrete step function. In effect, pixels darker than the threshold are placed in the FG layer. Since there are 64 pixels in a block, there are at most 64 different meaningful threshold values, whereby setting t_n to be beyond bounds forces the Mask block to be uniform, i.e. all samples imaged from one of the layers. Thresholding can be shown to be nearly RD-optimal under some circumstances. Please see [2],[9] for some details on thresholding analysis and its RD performance.

2.1. Finding the "best" threshold

If we sort the block pixels $x_n(i,j)$ into a sequence $p(k)$, for each $t_n = p(k)$, we evaluate

$$J_k = R_n(k) + \lambda D_n(k)\ , \tag{6}$$

where the index k denotes measurements for the k-th threshold tested as in (5). Both λ and Q_d are fixed for all image blocks because for optimality, blocks should operate at the same slope on their RD curves [10], and because baseline JPEG does not allow for changing quantizer tables within an image. We test all $p(k)$ in a block and select the index $k = k_o$ for the minimum J_k. Then, $m_n(i,j)$ is found using (5) for $t_n = p(k_o)$. An example is shown in Fig. 3:

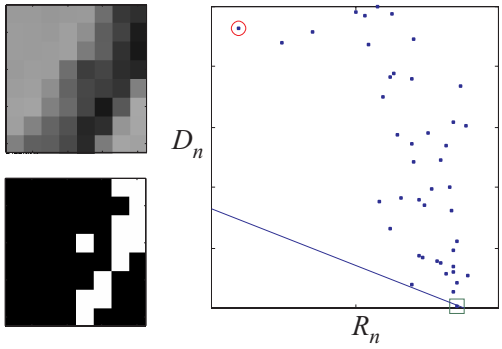

Figure 3. A sample blocks,its thresholding RD plot and resulting Mask block. (□ : best point); (◯: uniform mask).

2.2. Finding the block slope

In the previous section, we optimized the sequence $\{t_n\}$, block by block, for fixed external variables λ and Q_d. We want now to optimize these variables. As we discussed, in order to compute R and D, we have to decide upon Q_c. It is reasonable to assume $Q_d = Q_c = Q$ for a good RD estimation. In any case, R and D are functions of both λ and Q, i.e. $R(\lambda, Q)$ and $D(\lambda, Q)$. Given a budget R_b (or D_b), the goal is $\min_{\lambda,Q} D(\lambda,Q) | R(\lambda,Q) \le R_b$, i.e. we are interested in finding the lower convex hull (LCH) for a bounded RD region, preferably avoiding a search of the 2D (λ, Q) space.

We make the very reasonable assumptions that R and $1/D$ are monotonic increasing functions of both λ and $1/Q$. The higher λ is, the less importance R_n has in (6). Hence, the search will try to increase rate to reduce distortion which has larger weight. The higher Q, the coarser the quantization, thus the image incurs more distortion and is more easily compressed (reducing rate). We make yet another assumption. Let us start from an operational point $P_0 = (\lambda_0, Q_0)$. If we vary either Q or λ, which variation would cause a larger RD trade-off? We assume that Q does. The quantizer affects the RD trade-off directly. By changing λ, with Q fixed, the rate is changed only by making the Mask layer more or less active, which has only a modest impact on D. In other words,

$$\frac{D(P_0+\delta_\lambda) - D(P_0)}{R(P_0+\delta_\lambda) - R(P_0)} \le \frac{D(P_0+\delta_Q) - D(P_0)}{R(P_0+\delta_Q) - R(P_0)} \tag{7}$$

where $\delta_\lambda = (\Delta\lambda, 0)$ and $\delta_Q = (0, \Delta Q)$ are the minimum changes in order to produce a change in the RD point.

These assumptions imply that the mapping of the plane (λ, Q) into the plane (R, D) only warps the image into the range and the topology is maintained. The mapping is illustrated in Fig. 4, where the rectangular grid bounded by extreme values of λ and Q in Fig. 4(a) is mapped (distorted and stretched) into the irregular grid in Fig. 4(b) with the same topology (mapped lines do not cross). Thus, at most two of the boundary lines of the image in (λ, Q) are directly mapped to the LCH in the RD plane. However, λ has a limited effect on the RD point, while Q dominates the RD trade-off. As $Q \to \infty$, D diverges and R approaches its minimum (conceptually 0), regardless of λ. As $Q \to 0$, $D \to 0$ and R diverges, again, regardless of λ. Positive values of λ and Q, as illustrated in Fig. 4(c) are then mapped to a region as in Fig. 4(d). Thus, one of the axes is mapped to the LCH. Let $(R(P_0), D(P_0)) \equiv RD(P_0)$. Then, (7) is illustrated in Fig. 4(e). To confirm our speculations, tests were ran for different compound images for different rates and we verified

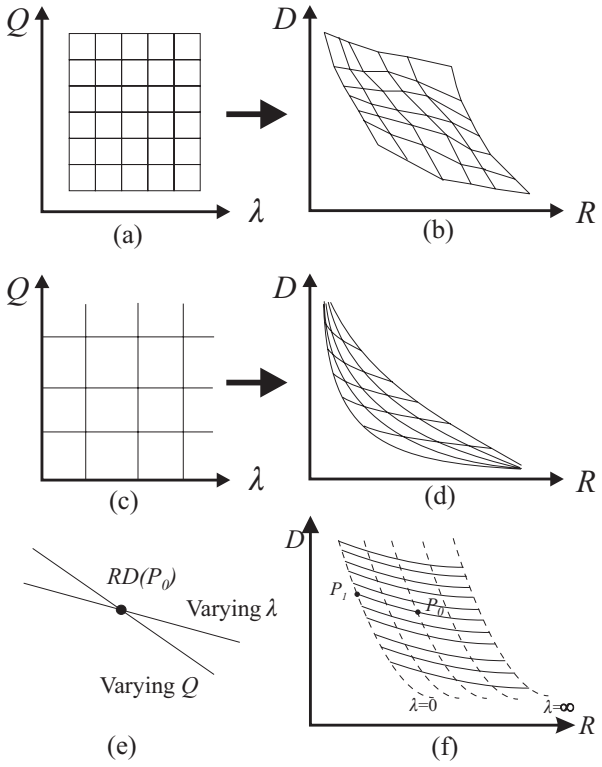

Figure 4. Mapping $(\lambda, Q) \rightarrow (R, D)$, variable dominance and determination of the LCH.

the relations that we hypothesized (monotonicity and (7)). In Fig. 5 it is shown RD curves by varying Q and by varying λ (minimum, intermediary, and maximum values). Circles mark the points where $Q = 1$.

The implications of this result are that if we start from point P_0 in Fig. 4(f), where the solid curves are those obtained by varying λ while the dashed lines are obtained by varying Q, it is easy to see that one will "hit" the LCH by moving leftward from point P_0 until getting to point P_1, i.e. the Q axis is mapped the LCH. In other words, one operates at the LCH if $\lambda = 0$! This unusual result translates in minimizing all R_n without regard to D_n, in order to minimize D for some R (or, R for some D).

2.3. Finding the "design" quantizer

Due to practical reasons which are discussed in [2], the otherwise logical concept of making $Q = Q_d = Q_c$ does not work very well for low rates. Large Q yields poor segmentation which somehow compromises the performance for moderate to high compression ratios. Note the erratic behaviour of curves in Fig. 5 for lower rates.

PSNR plots are shown in Fig. 6 for the image "compound1" from JPEG 2000's test set[1]. We compare the two algorithms in the MRC context. The solid curve represents the points obtained by optimizing for $Q = Q_d = Q_c$. The dashed curves correspond to fixing Q_d and varying Q_c. Note the poor performance of the first algorithm for lower bit rates which is only able to catch up with second algorithm's performance for high bit-rates and yet the gains are marginal. This effect is caused by the algorithm's inherent rate-estimation imprecision. Also, for high enough bit rates (low enough Q_d), the

[1]The set of images tested in this paper is shown in [2]

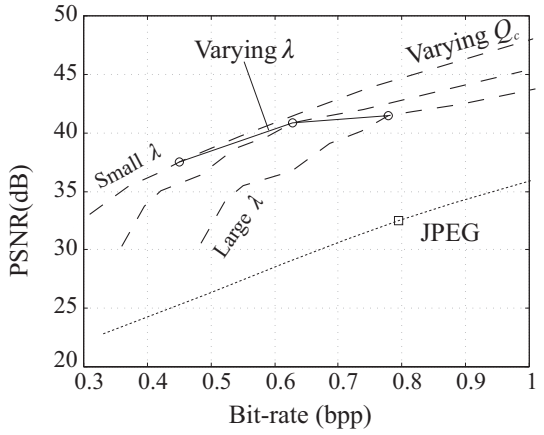

Figure 5. PSNR plots for MRC and JPEG using image "compound1", comparing . either varying λ or Q.

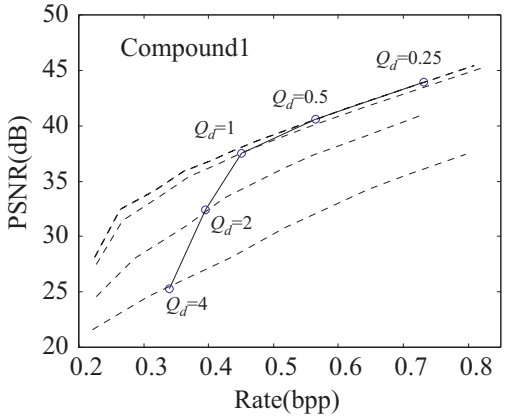

Figure 6. Comparison between varying Q_d (solid line) and Q_c only (dashed lines) for image compound1.

RD curves are virtually identical regardless of Q_d. This result was consistent with tests performed on other compound images. Summarizing, the preferred algorithm is the one where:

1. Select Q_d for high quality (e.g. $Q_d < 1$).
2. For block $x_n(i, j)$ and find $t_n | \min_{t_n} R(t_n, Q_d)$.
3. Obtain mask $m_n(i, j)$ for each block using (5).
4. Given $m(i, j)$, compress FG/BG layers for some Q_c.

The search in 2D space is avoided, i.e. the Mask layer can be generated with only one point in the (λ, Q) plane. Hence, a single RD-efficient segmentation algorithm can be made to approach or meet the LCH virtually independent of the RD slope, i.e. independent of the actual RD targets.

See [2] for a discussion on block rate estimation for high compression ratios and for computational complexity issues.

3. EXPERIMENTAL RESULTS

We now compare segmentation methods within our MRC framework using several images ranging from purely graphics (graphics) to purely pictorial (baby), with two other mixed images with graphics (compound1) and pictorial (wine) dominance. Clearly, the more graphics the higher is MRC's advantage over a single coder such as JPEG. For an image

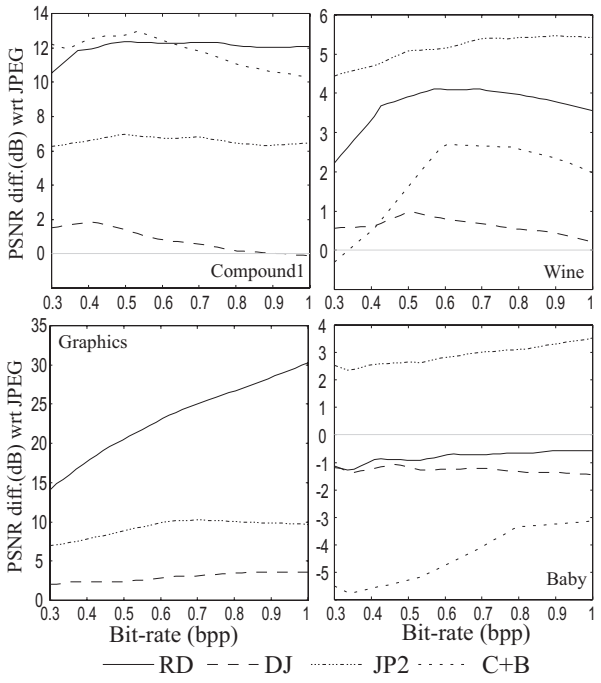

Figure 7. PSNR improcvements relative to JPEG. RD:proposed method,$Q_d = 1$; DJ and C+B are MRC with segmenters from [12],[11]; C+B is not available for image "graphics"; JP2 is JPEG 2000.

such as "baby", which is not supposed to be represented using MRC, there is the disadvantage of encoding the overhead of two planes and MRC is expected to be outperformed by JPEG. We compared the following segmenters: (i) $\lambda = 0$,$Q_d = 1$; (ii) the segmenter from [11]; and (iii) the segmenter from [12]. We also compared the single-layer coders JPEG and JPEG 2000 [13]. For the MRC approach, we computed RD curves by varying Q_c, shown in Fig. 7. The distortion measure chosen was PSNR and the plots present results in PSNR improvement compared to JPEG.

The PSNR difference against JPEG is extremely large for the graphics case since MRC quickly approaches the lossless state. The image compound1 is a typical target compound images, where thee is a staggering 12 dB improvement over JPEG (6dB over JPEG 2000). The performance of the variance-based method is very close to that of the RD-based one, except for pictorial images. As the image becomes purely pictorial, the losses are about or below 1 dB for the RD-based segmentation compared to JPEG. We consider this small loss a very positive sign: even if by mistake a pictorial image is to be segmented, smart segmentation can minimize losses.

The other segmenters tested do not employ block-thresholding as we do. The one from [11] is RD optimized for a blockwise multilevel mask which was adapted by the authors for our MRC approach, by thresholding blocks in the bimodal class. It slightly outperforms our RD-optimized approach for low bit-rates for image compound1, while being outperformed everywhere else. That result is a combination of their efficient bimodal classification with our deficient segmentation for low bit-rates. The multiresolution clustering segmenter from [12] was designed for high resolution images and is not very precise in identifying small text. Since it is intended for internet-based, very-high-compression applications, this is commonly satisfactory, but unfortunately yields modest PSNR performance for our small test images within our MRC approach. Both segmentation approaches serve to illustrate the potential gains one can obtain by optimizing the segmenter, as we propose, for a given MRC setup, instead of using a generic segmentation. See [2],[6] for examples of decompressed images, layers, etc.

4. CONCLUSIONS

Optimized block thresholding seems to be an effective way to segment a compound document image for compression. Because of the different curve slopes and of the practical estimation problems at low bit-rates, block rate minimization along with high-quality quantization seem to be the basis for a robust threshold-based segmentation method. The proposed segmentation method is RD-optimized in the sense of leading to points approaching or meeting the LCH, given all constraints, but is robust enough to be virtually independent of the RD slope and of the RD targets, which are basically set after segmentation, during the compression stage. Therefore, the segmenter can be applied to an image independent of the layer compression settings, which simplifies implementation. The variance method is a fast alternative, which can be used as an example where the RD-optimized method can guide the tuning of other segmenter's parameters.

Overall, the trade-off is reasonable: there can be improvement in the order of tens of dB in graphics case, while only a small loss in the pictorial case. Furthermore, it was shown a sizable improvement of the proposed methods over other MRC-centric segmenters.

REFERENCES

[1] R. Buckley, D. Venable and L. McIntyre, "New developments in color facsimile and internet fax," *Proc. IS&T's 5th Color Imaging Conference,* pp. 296–300, Nov. 1997.

[2] R. de Queiroz, Z. Fan and T. Tran, "Optimizing block-thresholding segmentation for multilayer compression of compound images," *IEEE Trans. Image Proc.,* Oct. 2000.

[3] Draft Recommendation T.44, Mixed Raster Content (MRC), ITU-T Study Group 8, Question 5, May 1997.

[4] R. de Queiroz, R. Buckley and M. Xu, "Mixed raster content (MRC) model for compound image compression," *Proc. VCIP,* SPIE Vol. 3653, pp. 1106-1117, Feb. 1999.

[5] IETF RFC 2301. File Format for Internet Fax. L. McIntyre, S. Zilles, R. Buckley, D. Venable, G. Parsons, J. Rafferty. March 1998. ftp://ftp.isi.edu/in-notes/rfc2301.txt.

[6] R. L. de Queiroz, "On data-filling algorithm for MRC," in this proceedings.

[7] W. P. Pennebaker and J. L. Mitchell, *JPEG: Still Image Compression Standard,* Van Nostrand-Reinhold, 1993.

[8] ITU-T Rec. T.6, Facsimile coding schemes and control functions for group 4 facsimile apparatus, Nov. 1988.

[9] R. de Queiroz, "Compression of compound documents," *Proc. ICIP,* 25PS1.1, Kobe, Japan, Oct. 1999.

[10] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression,* Kluwer Academic, 1992.

[11] H. Cheng and C. Bouman, "Document compression based on multiscale segmentation," *Proc. ICIP,* 25PS1.8, Kobe, Japan, Oct. 1999.

[12] L. Bottou, P. Haffner, P. Howard, P. Simard, Y. Bengio and Y. LeCun, "High quality document image compression using DjVu," *Journal of Electronic Imaging*, 7(3), pp. 410-425, July 1998.

[13] ISO/IEC JTC1/SC29 WG1, JPEG 2000 Committee, Working Draft 2.0, June 25, 1999.