HIERARCHICAL BILEVEL IMAGE COMPRESSION BASED ON CUTSET SAMPLING

Shengxin Zha[1], *Thrasyvoulos N. Pappas*[1], *David L. Neuhoff*[2]

[1]EECS Department, Northwestern Univ., Evanston, IL 60208
[2]EECS Department, Univ. of Michigan, Ann Arbor, MI 48109

ABSTRACT

We propose a hierarchical lossy bilevel image compression method that relies on adaptive cutset sampling (along lines of a rectangular grid with variable block size) and Markov Random Field based reconstruction. It is an efficient encoding scheme that preserves image structure by using a coarser grid in smooth areas of the image and a finer grid in areas with more detail. Experimental results demonstrate that the proposed method performs as well as or better than the fixed-grid approach, and outperforms other lossy bilevel compression methods in its rate-distortion performance.

***Index Terms*—** structurally lossless compression, rate-distortion, arithmetic coding, MRF

1. INTRODUCTION

Compression of bilevel images is important for a variety of applications involving text, graphics, halftones, pen and ink sketches, as well as for encoding segmentation information for image analysis and object-based compression (e.g., MPEG-4). The techniques we are concerned with in this paper apply to almost all bilevel images except halftones (that is, except images in which the black and white pixels are not organized in clusters). The JBIG standard (c.f. [1]) provides an efficient solution for the lossless bilevel compression case for all types of content, except for stochastic halftones. However, higher compression can be achieved by lossy techniques that can approximate the bilevel images with relatively small perceptual losses. The JBIG2 standard [2] aims to accomplish this with almost no degradation in image quality; it relies on partitioning the bilevel image into regions (of text, graphs, and halftones) and encoding each region with a different scheme. However, so far, it can only handle text and halftones. For more general bilevel images in which the black and white pixels are more or less organized into clusters, Reyes *et al.* [3] proposed a new method that relies on *cutset sampling* (along the lines of a rectangular grid) and Markov random field (MRF) constraints to reconstruct the smoothest image that is consistent with the cutset samples, which can be encoded in an efficient manner. They showed that their technique outperforms other lossy techniques [4].

In this paper we extend the cutset-MRF technique of [3], which losslessly encodes all pixels on a fixed rectangular grid and relies on the decoder to reconstruct the grid interiors, based on an MRF model. In the proposed approach we hierarchically adapt the grid size to local image detail. While this requires additional information for encoding the quadtree (decision bits for subdividing the grid and additional grid contours), we show that it provides a better overall rate-distortion performance by utilizing a coarser grid in smooth areas of the image and a finer grid in areas with more detail. The grid pixels and the side information are losslessly encoded using arithmetic coding with each pixel conditionally encoded given one previously encoded bit. We obtain accurate estimates of the encoding bitrate via the empirical first-order conditional entropy of the bits to be encoded.

Experimental results with a variety of bilevel images demonstrate that the proposed method offers an advantage over the fixed-grid approach. While the performance gains on most images are modest, our results contribute to a better understanding of the capabilities and limitations of cutset coding. Moreover, the method can be easily modified to obtain a progressive scheme, which starts with relatively smooth images and adds details as more bits are received.

In Section 2, we review the fixed-grid cutset-MRF approach. Section 3 describes the proposed adaptive technique, and Section 4 presents our experimental results.

2. REVIEW OF FIXED-GRID APPROACH

The encoder of the cutset-MRF technique of [3] uses a rectangular grid to subdivide the image into blocks (typically square), and then encodes the pixels on the block boundaries using a lossless compression technique, e.g., arithmetic or Huffman coding. The decoder losslessly reconstructs the pixels on the grid, and uses an MRF model to reconstruct the block interiors in a manner that obtains the smoothest bilevel image that is consistent with these grid pixels. Thus, the grid acts as a specification for the reconstruction by the decoder. An example is shown in Fig. 1. The role of the points on the grid is to preserve key structural information, on which the decoder bases the reconstruction. In this sense, it can be argued that, when the grid lines are not too far apart, this is an example of *structurally lossless* compression [5] for bilevel images, whereby the original and reconstructed images are similar and both of high quality, even though in a side-by-side comparison there may be visible differences.

The work of the third author was supported by NSF Grant CCF 0830438.

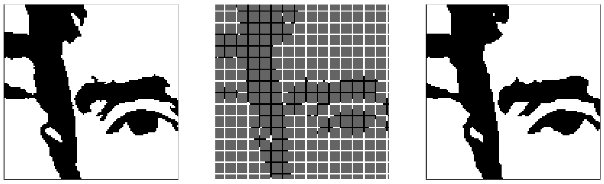

Fig. 1. Left to right: Original, specification, reconstruction

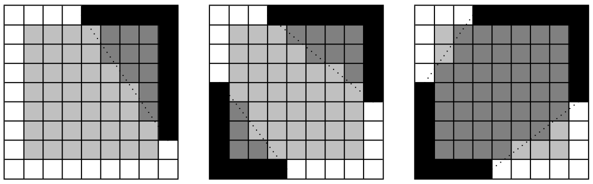

Fig. 2. Block Specifications and Reconstructions

The reconstruction in [3] is based on a MAP estimate of the interior pixels of a rectangular block, given the values of the boundary pixels, which reduces to finding the bilevel block interior that in combination with the boundary has the fewest black-white transitions between pixels. Thus, given the pixels on the boundary, each block is reconstructed independently of its neighbors. The authors found explicit rules for optimal reconstructions for several boundary specifications, that is, the boundary contains: (a) all black (or equivalently white) pixels; (b) one run of black (white) pixels; and (c) two runs of black (white) pixels. Since boundaries of three or more runs of black (white) pixels are not very common, they based the reconstruction in such cases on the two longest runs of one color, assuming all the other boundary pixels have the other color. In all of these cases, they found that the optimal reconstruction consists of monochrome regions bounded by (smooth, minimally varying) nonintersecting lines connecting the endpoints of the boundary runs. They also showed that, depending on the location of the endpoints, there may be a number of optimal reconstructions. In such cases, their decoder picks one at random. In this paper, for simplicity, we assume that the lines are straight; since such lines are included in the optimal set, there is no loss of optimality; moreover, this solves the ambiguity of multiple optimal solutions, and eliminates any biases towards black or white reconstructions. In the case of two black runs, there are four endpoints, and thus two possibilities: to connect the endpoints of each black run or to connect the endpoints of each white run. An optimal reconstruction can be found under each constraint. While only one of these reconstructions provides the overall optimum in the MAP sense, Reyes *et al.* [3] found it beneficial for the encoder to signify the one that best represents the original image via a *connection bit*, which is decided by comparing the two reconstructions to the actual block interior. Indeed, they showed that the connection bit contributes significantly to preserving structural information in the interior of the blocks. Examples of reconstructions are shown in Fig. 2. The gray pixels indicate pixel reconstructions, light gray as white and dark gray as black. Note that the two possible reconstructions for the boundary specification in

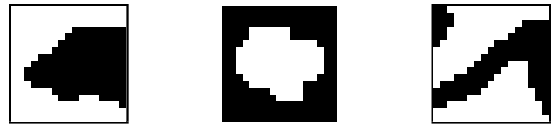

Fig. 3. Block Splitting

the middle and right. Note also the dotted lines that connect the endpoint of the boundary runs separating the black and white reconstructed pixels.

3. HIERARCHICAL APPROACH

We now consider the cutset-MRF encoding technique of the previous section with an adaptive grid. As in [3], the initial grid consists of every Nth row and column, indexed $(1+kN)$. This forms blocks of size $(N+1)\times(N+1)$ when the borders are included and $(N-1)\times(N-1)$ when they consist only of the interior pixels; adjacent blocks share a common border. In the following, for simplicity, we will refer to $N \times N$ blocks. The blocks are processed in raster order. Each block is reconstructed (at the encoder) from its border specification, as described in Section 2. The reconstruction is then compared with the original block using a similarity metric, and if the similarity is above a threshold, the block is subdivided into four subblocks, and the process is repeated until the threshold is met. For each block the encoder uses a *split bit* to indicate whether it was subdivided or not.

For each $N\times N$ block and subblock, the information to be encoded includes: (a) the top and left border of the block, (b) the internal grid pixels each time a block is subdivided, and (c) side information consisting of the split bit for each block and subblock, and the connection bit for blocks and subblocks with two or more runs on the boundary. As in [3], we considered the new approach with and without the connection bit, and found that the former result in considerably superior performance. The encoding of the grid pixels and side information can be done using a lossless arithmetic coding scheme, as discussed below.

The shape and size of black and white regions in a bilevel image is not necessarily homogeneous. The block splitting allows the technique to adapt the block size to local image detail. When the regions are large, a large block size is most efficient; when the regions are small, a small block size is needed to preserve image structure/detail. Fig. 3 shows examples of blocks whose interior cannot be adequately reconstructed based on the boundary pixels. Thus, splitting is necessary. Reyes *et al.* [6] investigated different metrics and criteria for connection decisions, and found that the fraction of pixels changed by the encoding/decoding process, which we call the *error rate*, works best. They also found that, while the error rate does not correlate well with image quality when comparing different compression approaches (e.g., [3] with [4]), when used within the MRF-based compression framework, it provides the most robust and perceptually meaningful results. Their conclusions are also relevant to the splitting decisions to be made in the proposed approach.

We considered adapting the splitting threshold to block size, but found that a fixed threshold works best. We also considered a metric that is equal to the maximum error rate over four or nine subblocks, rather than their average. Such a metric is intended to ensure a more even distribution of error around the block, especially when the block is large. Our results indicate that the particular choice does not make much difference in overall perceptual distortion (at a given encoding rate), even though it does have an effect on the distribution of errors around the image. Since there is no clearly winning strategy, we think that the choice should be left to the user.

As a last stage of the encoding process, one can eliminate border runs that are not connected to any interior pixels. Such runs can be due to rare blocks with more than two border runs, or isolated runs that are not connected to anything, as for example the one at the right side of the left block in Fig. 3. Even though the removal such runs increases the error rate, it improves perceptual quality (smoother image), and saves bits.

3.1. Lossless Coding

As described previously, encoding of an image is accomplished by losslessly encoding grid pixels, as well as connection and split bits. There are, of course, many potential ways to do this. Here, as in [3], we focus on arithmetic coding (AC) based on first-order conditional probabilities. Specifically, each of the three types of bits is arranged, i.e., ordered, into a separate stream to which AC is applied. AC requires that as each bit is fed to the encoder, it is accompanied by a probability distribution $\{p(0), p(1)\}$ that is relevant to this bit. (Actually, only $p(0)$ is needed.) In our tests of the new method, as in [3], $p(0)$ is an estimate of the conditional probability that the current bit is 0 given the value of a prespecified prior member of the stream, called its *context*.

For coding the grid pixels, the context of the current pixel is chosen to be another grid pixel, either the one just to its left or the one just above it. Thus, the grid pixel stream must be ordered so that every pixel in the stream is preceded by a left or upper neighbor. For example, one can scan the left and upper borders of each $N \times N$ block, taken in raster order, followed by the right and bottom borders of the entire image, and then the internal grid pixels. As the encoding proceeds, conditional probability estimates $p^n(0|0)$ and $p^n(0|1)$ are developed such that $p^n(0|i)$ is the probability input to the AC encoder when encoding the nth pixel in the stream and its context has value i. The estimate $p^n(0|i)$ is computed simply as the fraction of previous pixels that were 0 when their context was i. The resulting number of bits produced by AC is, with great accuracy, $-\sum_{n=1}^{M} \log_2 p^n(x_n|x_{c_n})$, where M is the total number of grid pixels, x_n is the nth grid pixel and c_n is the index of its context. When divided by M this can be viewed as an estimate of the first-order conditional entropy of the grid pixels. Since this formula is so accurate, we used it to determine coding rate, instead of explicitly applying AC.

We have also tried a *two-pass* AC and found, as typically happens, that *one-pass*, as described above, and two-pass yield essentially the same coding rate. Finally, AC with first-order conditional probability estimates was also used to encode the split and connection bits. However, for the latter, the gain over not encoding the bits was negligible.

4. EXPERIMENTAL RESULTS

The proposed method was tested on a variety of bilevel images and its performance was compared to the fixed-grid approach of [3]. Four of those images are shown in Figs. 4–8. The rate-distortion curves for the two techniques are also shown in the figures. The estimated rate of the one-pass first-order AC was used for these plots. The plots also show the performance of the method in [4] and JBIG, which is lossless. The points for the fixed-grid (no split) technique [3] were obtained by varying the grid size (1–16). The points for the hierarchical adaptive-grid technique were obtained by starting with block size 16, and varying the threshold for the block error rate. Starting with block size 16 yields better results than other starting sizes (32 or 8). The results in [4] were obtained by varying the *error rate factor.*

Fig. 4 shows an original 512×480 image and three reconstructions, corresponding to the points labeled in the curve; the associated encoding statistics are shown in Fig. 5: for each block size, fraction of image encoded with such, contribution to total bitrate, and fraction with connection bit, as well as, bitrate allocated to split bits. Note that the fidelity degrades gracefully with decreasing bitrate; one can argue that the image structure is preserved. At high quality levels the adaptive method offers a modest advantage over the fixed-grid approach, while as the rate (and quality) decreases, the performance in the rate-distortion sense is virtually identical. More importantly, the perceptual quality of decoded images with approximately the same error rate is similar. Actually, at the two highest rates the gain over the fixed-grid approach is substantial (25%), but we only claim a modest advantage because at these rates JBIG outperforms both techniques. As expected, the fraction of the image encoded with smaller blocks decreases with decreasing bitrate and so does the bitrate for the split bits, while the added bigger blocks use more connection bits. However, the contribution to total bitrate from the connection bits remains very small (.001 bpp).

The results for the images in Fig. 6 are comparable to those for the first image. On the other hand, Figs. 7 and 8 demonstrate a significant advantage of the proposed technique, both in terms of the rate-distortion curve, and in terms of the perceptual quality of images at approximately equal bitrates. Overall, we found that the performance advantage of the proposed technique varies from modest to substantial, depending on image content. We believe that the performance of the proposed approach will increase by a carefully designed second-order AC. More importantly, as we mentioned in the introduction, the proposed technique can easily be modified to make it *progressive*, by adding information at a finer grid size as additional bits are received.

Fig. 4. Hierarchical coding examples: Original, .09 bpp (A), .05 bpp (B), .04 bpp (C) (left to right), rate-distortion curves

	16	8	4	2	1	split	tot.
% image	72.7	15.6	7.7	3.7	.41		100
bitr. contr.	.048	.010	.005	.002	.0003	.021	.09
% con. bit	15.2	14.1	8.1	2.0	0		

	16	8	4	2	1	split	tot.
% image	85.2	11.5	2.9	.38	.06		100
bitr. contr.	.034	.005	.001	.0002	.00	.006	.05
% con. bit	20.6	23.8	16.8	3.25	0		

	16	8	4	2	1	split	tot.
% image	90.8	8.13	.96	.07	.02		100
bitr. contr.	.029	.003	.0003	.00	.00	.003	.04
% con. bit	23.7	26.7	25.3	0	0		

Fig. 5. Statistics for hierarchical coding examples A, B, and C in Fig. 4.

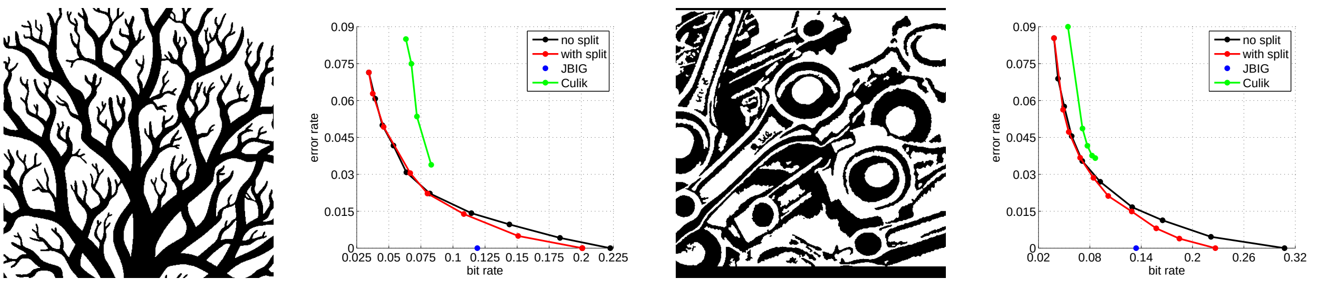

Fig. 6. Hierarchical Approach Examples: Original images and rate-distortion curves

Fig. 7. Fixed-Grid Coding: Original, 6×6 grid (0.052 bpp), 8×8 grid (0.04 bpp), 12×12 grid (0.03 bpp), (left to right)

Fig. 8. Rate-Distortion Curves; Hierarchical Coding: 0.052 bpp, 0.039 bpp, 0.03 bpp (left to right)

5. REFERENCES

[1] K. Sayood, *Introduction to Data Compression*, chapter 10, Morgan Kaufmann, 2006.

[2] P.G. Howard, *et al.*, "The emerging JBIG2 standard," *IEEE Tr. CSVT*, vol. 8, no. 7, pp. 838–848, Nov. 1998.

[3] M.G. Reyes, *et al.*, "Lossy compression of bilevel images based on Markov random fields," *ICIP-07*, pp. II–373.

[4] K. Culik II, V. Valenta, "Finite automata based compression of bi-level and simple color images," *Computer & Graphics*, vol. 21, pp. 61–68, 1997.

[5] T.N. Pappas, *et al.*, "Image analysis and compression: Renewed focus on texture," *Visual Infor. Proc. Commun.*, Jan. 2010, *Proc. SPIE*, Vol. 7543.

[6] M.G. Reyes, *et al.*, "Structure-preserving properties of bilevel image compression," *Human Vision Electr. Im. XIII*, Jan. 2008, *Proc. SPIE*, vol. 6806, pp. 680617–1–12.