

LOSSY COMPRESSION OF BILEVEL IMAGES BASED ON MARKOV RANDOM FIELDS

Matthew G. Reyes¹, Xiaonan Zhao², David L. Neuhoff¹, Thrasyvoulos N. Pappas²

¹EECS Department, Univ. of Michigan, Ann Arbor, MI 48109

²EECS Department, Northwestern Univ., Evanston, IL 60208

ABSTRACT

A new method for lossy compression of bilevel images based on Markov random fields (MRFs) is proposed. It preserves key structural information about the image, and then reconstructs the smoothest image that is consistent with this information. The smoother the original image, the lower the required bit rate, and conversely, the lower the bit rate, the smoother the approximation provided by the decoded image. The main idea is that as long as the key structural information is preserved, then any smooth contours consistent with this information will provide an acceptable reconstructed image. The use of MRFs in the decoding stage is the key to efficient compression. Experimental results demonstrate that the new technique outperforms existing lossy compression techniques, and provides substantially lower rates than lossless techniques (JBIG) with little loss in perceived image quality.

Index Terms— Rate-distortion, structural coherence.

1. INTRODUCTION

Compression of bilevel images is important by itself, and also for coding of segment information in object-based coding (e.g., MPEG-4 [1]) and second generation image coding techniques [2]. However, it is well-established in the literature that the bit rate required for contour coding is quite high, and accounts for a significant percentage of the overall bit rate of such techniques. The JBIG standard (c.f. [3]) provides an efficient solution to the lossless bilevel compression problem for a variety of image types that include text, graphs, binary objects of various shapes, and even some halftoning techniques. However, the resulting bit rates are still relatively high, and there is a need for lossy techniques that provide high fidelity approximations of the bilevel images without substantial perceptual losses. The recently proposed JBIG2 standard [4] aims to provide much higher compression ratios with almost no degradation in image quality. In JBIG2, the bilevel image is typically divided into regions (text, graphs, halftones), and each region is encoded with a different scheme. The first stage of the standard deals with bilevel images of text and halftones. In addition, there is a substantial literature on lossy compression techniques for coding object contours (e.g., for MPEG-4 [5], [6]). The applicability of such techniques is somewhat limited, however, because they assume well-defined binary

objects. A number of other lossy techniques have been proposed in the literature (c.f. [7, 8]), but their rate-distortion performance is not as competitive. In this paper, we propose a new scheme for lossy coding of bilevel images that uses Markov random fields (MRFs) to reconstruct the smoothest image that is consistent with key structural information that can be encoded in an efficient manner. See Fig. 2 for examples of images to be considered.

The goal of lossy bilevel image compression is to provide a balance between bit rate and reconstructed image quality. We measure image quality by the fraction of pixels changed by encoding/decoding, which we call the *error rate*. (For bilevel images, this is equivalent to mean squared error.) The main idea of the proposed approach is that as long as the main structural information in the image is preserved, the details of the reconstructed image have very little effect on perceived quality. Thus, our goal is to reconstruct the smoothest image that is consistent with the transmitted structural information. For the encoding of the structural information we divide the image into blocks, and losslessly encode the pixels on the block boundaries. We then use MRF properties to reconstruct the smoothest bilevel image that is consistent with this specification. We consider a fixed block size for the specification, but the proposed method can be easily extended to a quadtree approach that adjusts the block size to local image detail.

The theory of Markov random fields [9, 10] has been used extensively in image processing, especially in segmentation, texture analysis and synthesis, restoration, and other applications, including image compression [11]. One of the attractive properties of MRFs is their ability to impose some degree of structural coherence (spatial continuity constraints) to an image or image region. While the use of MRFs in image compression is not new, the proposed approach is unique in using MRFs solely for image reconstruction in the decoder.

In the remainder of this paper, Section 2 provides an overview of the encoding system and details of the encoder; Section 3 describes the MRF model we use; Section 4 describes the reconstruction algorithm used at the decoder, and Section 6 discusses results and conclusions.

2. SYSTEM OVERVIEW

In the proposed compression system, the encoder losslessly encodes a prespecified subset of the image pixels, and the de-

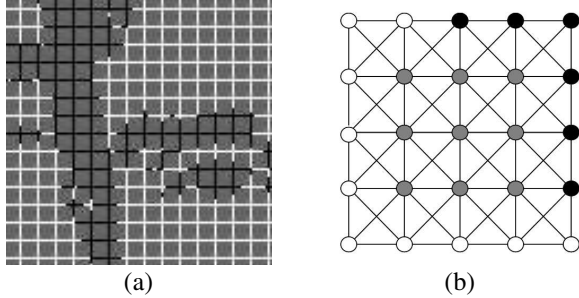


Fig. 1. (a) Detail of specification using $N = 8$ grid. The interior pixels of each block are shown in gray. (b) The nodes and edges between nodes of a 5×5 block whose boundary contains one run of consecutive 1s, with endpoints on different sides. Interior pixels are shown in gray.

coder performs (primarily) maximum a posteriori (MAP) reconstruction of the remaining pixels (given the encoded subset) with respect to a Markov random field model.

For the encoding, we fix a block size N , typically ranging from 2 to 16, and losslessly encode the pixels in an $N \times N$ square grid. That is, we losslessly encode only those pixels in rows $1, N+1, 2N+1, \dots$, or in columns $1, N+1, 2N+1, \dots$. We also losslessly encode the right column and bottom row, if they were not already encoded. Figure 1 shows a portion of the pixels in an $N = 8$ grid.

There are many potential ways to losslessly encode the pixels in the grid. For example, they might be arranged into a sequence for arithmetic or runlength coding. However, in this paper, as a simple demonstration of the potential of the new method, we estimate encoded bit rate by computing the empirical conditional entropy $H(X_2|X_1)$ of a grid pixel given a horizontally or vertically adjacent pixel, which is an accurate estimate for arithmetic coding based on first-order conditional probabilities. Higher-order conditioning would work even better.

As discussed later, the MRF model will be such that the interior of each grid block can be MAP reconstructed based only on the boundary of that particular block. Simple rules for MAP reconstruction are discussed in Section 4.

3. MARKOV RANDOM FIELD MODEL

Recall that an MRF is specified by a graph $G = (V, E)$ and a collection of *clique potential functions* $\Psi_c(x)$. Here V is a set of *nodes* – in our case the pixel locations of the image – and E is a set of pairs of nodes. The members of E are considered to be (undirected) *edges* of the graph. Moreover, E defines a *neighborhood relation* between nodes, in the sense that two nodes are neighbors if and only if they are connected by an edge. A *clique* c is a subset of nodes in V such that all pairs of nodes in c are neighbors, i.e. are connected by an edge. Let C denote the collection of all cliques in G . Now consider a bilevel image x that assigns a zero (white) or one (black) to each pixel location in V . For each $c \in C$, a clique potential function $\Psi_c(x)$ assigns a value to x that depends only the

pixel values of x in the locations specified by c . Now, for this MRF model, the probability of realization x is given by $p(x) = \frac{1}{z} \exp \left\{ - \sum_{c \in C} \Psi_c(x) \right\}$, where z is a normalizing constant. MRFs have the key property that the probability of a realization over a subset of nodes, given a realization over the neighbors of the nodes, is independent of the realization on the rest of the nodes.

We choose the MRF model to have edges that connect horizontally, vertically and diagonally adjacent nodes. Accordingly, the cliques are singletons, pairs of adjacent nodes, triples of adjacent nodes forming a right triangle, and quadruples forming a square. For cliques consisting of a single pair of nodes, the potential function assigns -1 if the nodes have the same value under x , and assigns +1 if they have different values. For cliques consisting of one, three or four nodes, the potential function always assigns 0. Under this model, homogeneous regions and continuous contours are favored over noisy regions and discontinuous contours.

4. DECODING/RECONSTRUCTION

The decoder begins by losslessly decoding the encoded grid. Then it visits each block in raster order and computes (primarily) a MAP estimate of its interior given the values on its boundary. Note that under the assumed MRF model, the interior of a block can be reconstructed using only the pixels on the boundary of that block. Moreover, the conditional probability of a block decreases monotonically with the number of pairs of adjacent nodes in the block having different values, which we call *dissimilar pairs*. Therefore, a block has maximum probability given the boundary if and only if it has fewest dissimilar pairs.

There are several cases to consider. First, if the boundary consists entirely of 0s (respectively 1s), then it is easy to see that the MAP reconstruction is entirely 0s (respectively 1's).

Second, suppose the boundary contains one run of consecutive 1s, and all other boundary pixels are 0. Denote the first and last pixels of the run by P1 and P2. If P1 and P2 are on the same side of the block, then it is easy to see that the MAP reconstruction fills the interior with all 1's or all 0's, depending on which is more prevalent on the boundary. If P1 and P2 are on different sides, as illustrated in Fig. 1, then it is easy to show that a MAP reconstruction is determined by a *path* of pixels from P1 to P2 through the interior of the block. (A path is a sequence of distinct neighboring pixels.) The pixels on the path, and those between the path and the run of 1s on the boundary, are reconstructed as 1; all others as 0. It can be shown that if the vertical distance between P1 and P2 exceeds the horizontal distance, then a path is optimal, i.e., it determines a MAP reconstruction, if and only if it consists entirely of vertical and diagonal steps in the direction of P2. A similar rule applies when the horizontal distance exceeds the vertical distance. It is now evident that there are usually a number of MAP reconstructions. In the results in Sec. 6, we select one at random.

Third, if the boundary contains two runs of 1s separated by zeros, then it can be shown that the MAP reconstruction is obtained by selecting two paths, each connecting a pair of run endpoints, and filling in the remaining pixels in the natural way. There are two potentially optimal ways of pairing endpoints. For any given pairing, a path is optimal if and only if it satisfies the condition in the second case above. Each way of pairing run endpoints can be considered and the one creating a block with fewest dissimilar pairs is selected.

Finally, when the boundary contains three or more runs of 1s, true MAP reconstruction becomes much more complex, especially when N is not small. However, since this does not happen very often, we simply use the ad hoc approach of choosing the two longest runs of 1s on the boundary, and reconstructing the interior as described in the previous paragraph, as if all other pixels on the boundary were 0.

5. ENCODING WITH A DECISION BIT

As will be evident in Sec. 6, when a block has two runs of 0's and 1's on its boundary, the MAP reconstruction sometimes chooses an unfortunate pair of paths. For this reason, we have found it useful to modify the encoding algorithm so that when there are two runs on the boundary, both ways of pairing run end points are tried, and one bit is used to indicate that which creates the reconstruction that best matches the original. A similar strategy is used when there are three or more runs on the boundary. It will be seen that this results in a significant increase in reconstruction quality for a small increase in bit rate.

6. RESULTS AND DISCUSSION

To test the proposed method, we ran it on two sets of bilevel images, each set representing a scene with three different levels of smoothness. We used the adaptive clustering algorithm of [12] on two original grayscale images to generate bilevel images with different degrees of smoothness by controlling the algorithm parameters. This way we can control for image content and evaluate the response of our system to varying degrees of smoothness. The roughest and smoothest of each set of three are shown in Fig. 2. All images are 512x512.

In addition, we tested the lossy bilevel coding scheme of Culik and Valenta [7,8] on the same images, as well as a lossy coder consisting of a morphological (median-type) smoothing filter followed by lossless JBIG compression.

Figure 3 shows error rate vs. bit rate for each of the six test images and the various encoding methods. For our methods the different points in each curve were obtained by choosing block sizes 16, 14, 12, 10, 8, 6, 4, 2, 1, which have decreasing error rate and increasing bit rate. For the Culik-Valenta method, the points were obtained by varying their quality factor. For the smoothing plus JBIG the points were obtained by varying the window size of a morphological filter. Each plot also includes the rate of lossless coding with our method ($N = 1$) and with JBIG.

We see from these plots that the decision-bit MRF method is significantly better than both the basic MRF method and the Culik-Valenta method, with larger improvements at lower rates. We also see that it is much better than the JBIG smoothing method, except at very low error rates. For instance, at zero error-rate, JBIG is better than our $N = 1$ lossless coder, which is to be expected, since it is overly simplistic.

Figure 4 shows, again, the leftmost image of Fig. 2, as well as the results of encoding/decoding with, respectively, the basic method of Sections 2 and 4 ($N = 8$), the decision-bit method of Sec. 5 ($N = 8$), and the method of Culik-Valenta at a similar error rate. Note that the basic method has turned some thin black lines into dotted lines (for example in the shirt collar and the lines above the head), and the eyes have been largely filled in with black. This is due to the MAP reconstruction of blocks whose boundaries have two or more runs of 1s. The reconstruction in the decision-bit method does not suffer from this. We conclude that the benefits of the decision-bit method in preserving the structural integrity of the image are larger than what one might expect from the error rate improvements shown in Fig. 3. In comparison, the Culik-Valenta method tends to produce reconstructions with rough contours and unnaturally sharp angles. In summary, we have observed that the decision-bit MRF method generally produces reconstructed images that preserve structure and overall quality at favorable bit rates.

7. REFERENCES

- [1] ISO/IEC 14496, "Information technology – generic coding of audio-visual objects," July 1999.
- [2] M. Kunt, A. Ikonomopoulos, M. Kocher, "Second-generation image-coding techniques," *Proc. IEEE*, vol. 73, no. 4, pp. 549–574, Apr. 1985.
- [3] K. Sayood, *Introduction to Data Compression*, chapter 10, Morgan Kaufmann, 2006.
- [4] P. G. Howard, *et al.*, "The emerging JBIG2 standard," *IEEE Tr. Circ. Syst. Video Tech.*, vol. 8, no. 7, pp. 838–848, Nov. 1998.
- [5] A. K. Katsaggelos, *et al.*, "MPEG-4 and rate-distortion-based shape-coding techniques," *Proc. IEEE*, vol. 86, no. 6, pp. 1126–1154, June 1998.
- [6] A. Vetro, Y. Wang, H. Sun, "Rate-distortion modeling for multiscale binary shape coding based on Markov random fields," *IEEE Tr. Im. Proc.*, vol. 12, pp. 356–364, Mar. 2003.
- [7] K. Culik II, V. Valenta, "Finite automata based compression of bi-level and simple color images," *Proc. IEEE Data Compression Conf.*, Snowbird, Utah, 1996.
- [8] K. Culik II, V. Valenta, "Finite automata based compression of bi-level and simple color images," *Computer & Graphics*, vol. 21, no. 1, pp. 61–68, 1997.
- [9] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Royal Stat. Soc. B*, vol. 26, no. 2, pp. 192–236, 1974.

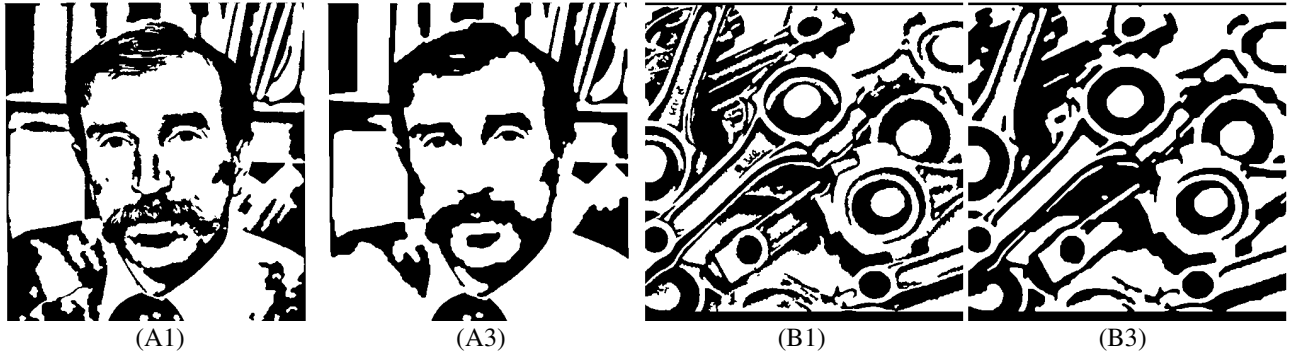


Fig. 2. The roughest and smoothest version of each test image.

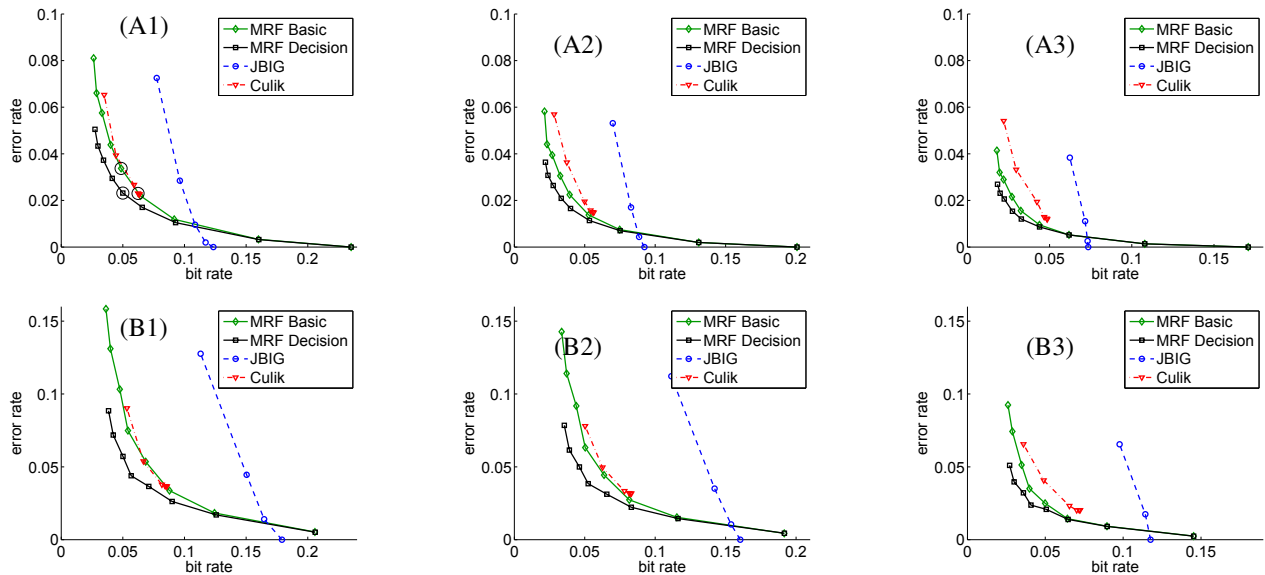


Fig. 3. Error rate vs. bit rate for the various encoding methods and three versions of the two test images. The decoded images corresponding to the circled points are shown in Fig. 4.

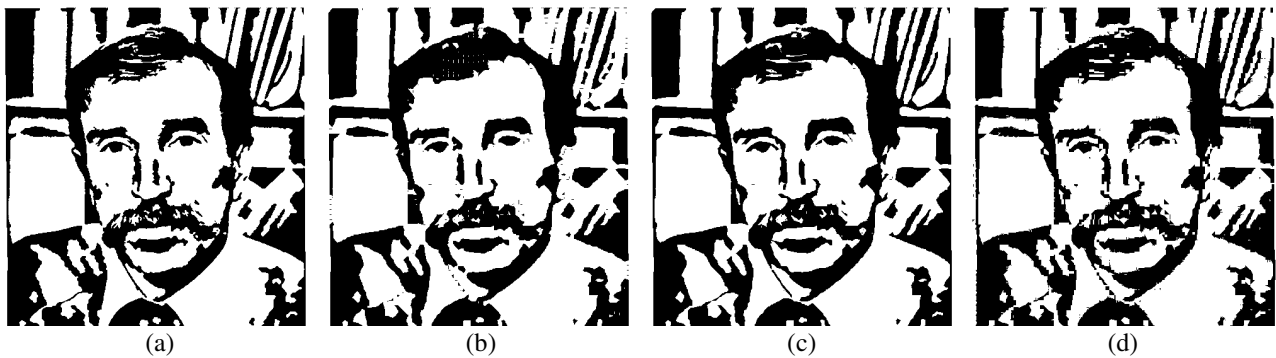


Fig. 4. (a) The leftmost image of Fig. 2. Its reproduction after being coded/decoded with (b) the basic method ($N = 8$), (c) the decision-bit method ($N = 8$), and (d) the Culik-Valenta method at a similar error rate.

[10] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.
 [11] G. Gelli, G. Poggi, R. P. Ragozini, "Multispectral-image compression based on tree-structured Markov random filed segmentation and transform coding," *Proc. IEEE*

Geosc. and Rem. Sens. Symp., 1999, vol. 2, pp. 1167–1170.
 [12] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Tr. Signal Process.*, vol. SP-40, no. 4, pp. 901–914, Apr. 1992.