

SIMPLIFIED GRID MESSAGE-PASSING ALGORITHM WITH APPLICATION TO DIGITAL IMAGE HALFTONING

Phunsak Thienviboon, Antonio Ortega, and Keith M. Chugg

Dept. of Electrical Engineering-Systems
University of Southern California
thienvi@usc.edu, ortega@sipi.usc.edu, chugg@usc.edu

ABSTRACT

Based on message-passing techniques, a novel iterative grid algorithm for the general two-dimensional (2D) digital least metric (DLM) problem is proposed and applied to image halftoning. The algorithm attempts to achieve a globally optimal solution via a local-metric computation and message passing as opposed to other 2D iterative global-metric optimizations such as simulated annealing and toggle/swap scheme. A reduced-complexity version of the proposed digital image halftoning technique is demonstrated. Results show that the quality of the halftone images is comparable to that of the state-of-the-art toggle/swap algorithm. Since the algorithm is not constrained by the specific metric used in this work, the proposed method is directly applicable to other digital image processing tasks (*e.g.*, optimal near-lossless coding or entropy-constrained halftoning). An expanded set of results from this work can be viewed at <http://sipi.usc.edu/~ortega/icip2001/icip2001.html>

1. INTRODUCTION

General two-dimensional (2D) digital least-metric (DLM) problems [3] can be found in a broad range of applications in digital imaging (*e.g.*, processing, compression, and generation), page-oriented memories (POM), and concatenated systems in digital communications. However, unlike in their 1D counterparts, there are no efficient approaches to conduct an exhaustive search for 2D DLM problems. Common approaches to tackle the 2D DLM problems include the application of 1D DLM algorithms (*e.g.*, Viterbi algorithm (VA), ML algorithm, etc.) or 2D heuristic optimizations based on a global metric (*e.g.*, simulated annealing, toggle/swap algorithm). Recently, after the decoding algorithm of turbo codes was introduced, iterative message-passing algorithms have proven to be a powerful tool for many DLM problems in the communications literature [2, 5]. Such algorithms utilize a local message computation and propagation in order to reach near optimal solutions after several iterations. These message-passing approaches are variously

known as belief propagation [10], generalized distributive law (GDL) [1] and sum-product algorithms (in the factor graph research area [5].)

In this paper we apply message-passing techniques to the 2D DLM problem arising in image halftoning, a process that creates an illusion of a continuous-tone image with a black-and-white image. Numerous authors have proposed to address this problem as a 2D DLM problem and have proposed solutions based on algorithms such as simulated annealing [4], neural networks [4], the Viterbi algorithm [8], multipath tree-coding [12], and the toggle/swap algorithms [6, 7, 9]. In [3], the iterative message passing algorithm is introduced to the halftoning problem with the row/column processing. However, the algorithm presented here achieves better performance with less complexity than the algorithm in [3]. The first novelty in our work is the introduction of a loopy graphical model, which we call the *grid model*, that enables the application of an iterative message-passing technique to the 2D DLM problem. We also introduce reduced-complexity techniques that prove useful for the digital image halftoning problem but could be used in other scenarios as well. The results show that the halftone image quality we achieve is comparable to that of the state-of-the-art toggle/swap scheme. Note that while our results are for a halftoning problem, the general framework we propose to solve the 2D DLM problem can be applied to other problems in digital imaging (*e.g.*, optimal near-lossless compression [3] or entropy-constrained optimal halftoning [12]) with appropriate changes (*e.g.*, proper metric definition).

In the next section, the iterative message-passing algorithm for 2D DLM is described. The reduced complexity method for digital image halftoning is proposed in Section 3. Section 4 illustrates the results and provides a discussion of the algorithm.

2. GRID ALGORITHM

The 2D DLM problem [3] is a problem of finding the intensity level for each pixel $b_{i,j} \in \mathcal{B}$ (finite) that minimizes an additive cost metric with local dependencies $\lambda(\cdot)$. Specifi-

cally, the desired image is $\hat{\mathbf{B}}$ which satisfies

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lambda(z_{i,j}; T_{i,j}) \quad (1)$$

where $\mathbf{B} = \{b_{i,j}\}_{i,j=0}^{N-1}$ is a candidate image, $T_{i,j}$ is the value of this candidate image in the neighborhood of (i, j) , and $z_{i,j}$ is the observation at location (i, j) . Particularly, $z_{i,j}$ is the (filtered) original gray-scale intensity at location (i, j) in the halftoning problem and $z_{i,j}$ is the received intensity at location (i, j) for the image deblurring problem. In addition, the support region of $\lambda(z_{i,j}; T_{i,j})$ determines the number of possible values of $T_{i,j}$ – e.g., if the support region is 3×3 and $b_{i,j}$ is binary, there are 2^9 possible values of $T_{i,j}$, which we refer to as *local configurations*.

The 2D DLM problem may be written as a two-step problem of the form

$$\hat{b}_{k,l} = \arg \min_{b_{k,l}} \left[\min_{\mathbf{B}} \sum_{i,j=0}^{N-1} \lambda(z_{i,j}; T_{i,j}) \right] \quad (2)$$

where the *inner* minimization is conducted for all image candidates corresponding to the conditional value $b_{k,l}$. The problem of computing this inner minimization in (2) is a well-known problem in engineering and computer science (e.g., see [1, 2, 5] and references therein). Algorithms based on the notion of message-passing on graphical models may provide efficient solutions to this problem. Specifically, if the underlying graphical formulation is cycle-free, message-passing algorithms provide the desired solution. Recently, however, it has been widely appreciated that message-passing algorithms can be very effective in practice even when the underlying graphical model contains cycles (i.e., when there is no theoretical assurance of optimality). In the following we describe briefly the operation of a message passing algorithm based on a novel graphical model for the 2D DLM problem in (2). The underlying principles used to define the standard messages and update rules have been described from a variety of perspectives and the reader is referred to the following for details: [1, 2, 5, 10].

The underlying graphical model used is the *grid model* shown in Fig. 1(a). The two types of nodes represent the image variables $b_{i,j}$ and the local configurations $T_{i,j}$, respectively. The edges connecting the configuration nodes represent the “overlap” (or *mutual information*) between the configurations that they connect. Applying the standard message passing technique, there is a processing node corresponding to each configuration node in Fig. 1(a). The role of this node is to accept some messages from the other (directly connected) processing nodes, and to return a similar message to each of those nodes. These messages represent some measure of the quality of each of the possible configurations of the edge variables. By repeated activation of

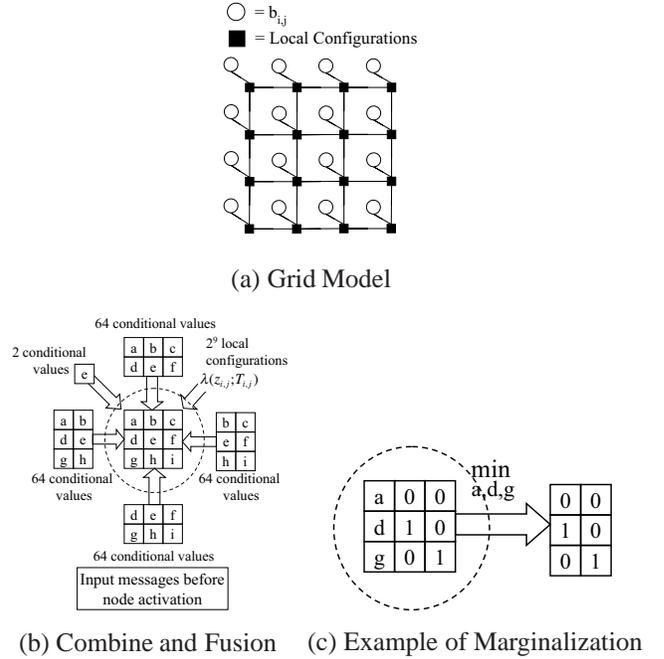


Fig. 1. Graphical Model for Grid Algorithm.

these processing nodes, the global configuration (i.e., \mathbf{B}) is eventually accounted for via these edge messages.

For a concrete example, consider a 3×3 support region and binary $b_{i,j}$ and the node processor at location (i, j) . Each of the four edge variables comprises 6 binary pixels, so that each takes on 64 conditional values. Let $U_{i,j}$, $D_{i,j}$, $L_{i,j}$, and $R_{i,j}$ define the edge variables in the “up”, “down”, “left” and “right” directions from (i, j) , respectively. For a given variable v taking a finite number of values, let $\text{MI}[v]$ denote the incoming messages to the node processor at (i, j) , then for each of the 2^9 possible local configurations, we compute the metric

$$\begin{aligned} \text{M}[T_{i,j}] &= \lambda(z_{i,j}; T_{i,j}) + \text{MI}[b_{i,j}] \\ &\quad \text{MI}[U_{i,j}] + \text{MI}[D_{i,j}] + \text{MI}[L_{i,j}] + \text{MI}[R_{i,j}] \end{aligned} \quad (3)$$

Note that each of the conditional variables on the right-hand side of (3) are determined uniquely by the configuration $T_{i,j}$ considered. This message combining or fusing process is illustrated in Fig. 1(b). The messages returned in each direction are obtained by minimizing over consistent configurations and subtracting the corresponding input message. For example, the message sent out to the right is

$$\text{MO}[R_{i,j}] = \min_{T_{i,j}, R_{i,j}} \text{M}[T_{i,j}] - \text{MI}[R_{i,j}] \quad (4)$$

where the minimization is conducted for all local configurations $T_{i,j}$ corresponding to the edge variable $R_{i,j}$. This marginalization processing is shown in Fig. 1(c). The messages sent out in other directions are defined analogously.

An algorithm for the 2D DLM problem is then obtained by specifying a schedule for activating the node processors and a stopping condition. After stopping the algorithm, the value of the halftone image at location (i, j) is selected to be the minimizer of the quantity $\text{MO}[b_{i,j}] + \text{MI}[b_{i,j}]$.

3. REDUCED-COMPLEXITY GRID ALGORITHM FOR DIGITAL IMAGE HALFTONING

In this paper, we consider the least-squares optimization corresponding to a halftoning problem, defined as

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (z_{i,j} - \mathcal{X}(T_{i,j}))^2 \quad (5)$$

where $b_{i,j}$ is the intensity level of halftone image, $z_{i,j}$ and $\mathcal{X}(T_{i,j})$ are filtered versions of gray-scale and halftone images, respectively, $\mathcal{X}(T_{i,j}) = b_{i,j} * h_{i,j}$ and $z_{i,j} = y_{i,j} * h'_{i,j}$, where $y_{i,j}$ is the original gray-scale image. The filters $h_{i,j}$ and $h'_{i,j}$ are 2D eye filters. All images are $N \times N$ and black (white) pixels are represented by 1 (0) intensity level. Therefore, the grid algorithm is applicable with, however, an exponential complexity of the size of $T_{i,j}$ or the size of the support for the filter h . This requires a number of reduced-complexity techniques to enable the use of sufficiently large filters.

To limit the complexity of message updating, let $T'_{i,j}$ be a L-shape pattern $\{b_{i-1,j}, b_{i,j-1}, b_{i,j}\}$ with 8 possible configurations. The square-error metric is modified to be

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (\tilde{z}_{i,j} - \mathcal{X}'(T'_{i,j}))^2 \quad (6)$$

by assigning the most updated decision of the intensity levels, $\hat{b}_{i,j}$, to all pixels in $T_{i,j}$ but outside $T'_{i,j}$. Therefore,

$$\mathcal{X}'(T_{i,j}) = h_{1,0}b_{i-1,j} + h_{0,1}b_{i,j-1} + h_{0,0}b_{i,j} \quad (7)$$

$$\tilde{z}_{i,j} = z_{i,j} - \check{h}_{i,j} * \hat{b}_{i,j} \quad (8)$$

$$\hat{b}_{i,j} = \arg \min_{b_{i,j}} \text{MO}[b_{i,j}] + \text{MI}[b_{i,j}] \quad (9)$$

where the filter $\check{h}_{i,j}$ is an eye filter $h_{i,j}$ with $h_{1,0} = h_{0,1} = h_{0,0} = 0$ and $\hat{b}_{i,j}$ is the current decision value of $b_{i,j}$. Note that, $U_{i,j} = b_{i-1,j}$, $L_{i,j} = b_{i,j-1}$, and $D_{i,j} = R_{i,j} = b_{i,j}$. Define a node activation as a operation of message updating (i.e., (3) and (4)), thresholding for bit decision (9), and decision feedback (8). Then, the reduced-complexity algorithm is conducted by running a sequence of node activations as follows. For each iteration, nodes $T_{i,j}$ are activated row-wise left-to-right and right-to-left from top-to-bottom, and then column-wise top-to-bottom and bottom-to-top from left-to-right. This will be run iteratively until a stopping condition is reached. For numerical stability, a

algorithm	square-error cost metric (per pixel)
Reduced-Complexity Grid algorithm	0.000260
Toggle/Swap [6, 7, 9]	0.000139
Toggle Only [6, 7, 9]	0.000417
Floyd-Steinberg error diffusion [11]	0.000347

Table 1. Comparison of square-error cost metric.

metric normalization is done so that the metric corresponding to 0-bit is 0. This allows all messages are represented by one number and simplifies the additions in (3) and (4). Therefore, the message updating requires 8 multiplications, at most 30 additions, and at most six 4-way compare/select operations. The decision feedback for each $\hat{b}_{i,j}$ requires a $|T_{i,j}|-3$ additions. Each node is activated 4 times in each iteration. Note that the complexity of the associated look-up-table (LUT) is considered negligible.

4. RESULTS AND DISCUSSION

To test the reduced-complexity algorithm in Section 3, the 512×512 gray-level Lenna image was used as a test image. The filters $h_{i,j}$ and $h'_{i,j}$ are Gaussian with $\sigma = 1.5$ (truncated to 9×9) and $\sigma = 0.9$ (truncated to 5×5), respectively. The initial binary image is randomly generated and all messages are initiated with zero metric. Fig. 2 shows a result of this algorithm at the 10th iteration and Table 1 provides a comparison in terms square-error cost metrics among various techniques. It was found that these metrics can roughly characterize the subjective quality of the halftone images. Using the same metric, the image from a toggle/swap technique (after 19 iterations) is shown in Fig. 3. As a result, the image quality from the proposed technique is comparable to that of the toggle/swap technique (used in 2D LSMB [9] and DBS [6, 7]) which is regarded as a very high-quality halftoning process.

In the comparison with the toggle/swap scheme, the specific DBS approach we simulated is that in [7], which, to the best of our knowledge, is the highest performance one to date. When comparing the relative complexities of the grid technique and the DBS approach we implemented both without filter- or image-dependent optimizations. Thus we do not use any initialization in either technique and we do not explore potential complexity reductions that may arise from symmetries in the values of the filter coefficients. In fact, for the DBS approach, this is called the *original version* in [7]. Under these conditions the complexity of the two approaches is comparable, with the grid technique being generally faster than toggle/swap and about the same speed as toggle-only schemes. Note that this complexity does not include the complexity of the table initialization for both techniques. In [7], a number of techniques have



Fig. 2. Reduced-complexity Grid algorithm (10th iteration).



Fig. 3. Toggle/Swap at convergence.

been introduced to accelerate the DBS operation. Most of these techniques (*e.g.*, choosing a good initial image before the start of the iteration) can be used within our framework as well so that we expect our algorithm, once filter- and image-dependent optimizations have been incorporated, to operate at least at a comparable speed. In addition, other techniques can be used to reduce complexity of the grid algorithm. In fact, it is straightforward to show that the error diffusion [11] and 1D LSMB [8] can be derived from this grid algorithm but with the recursive and 1D eye filters, respectively. Finally, it is worth noting that the algorithm introduced herein is applicable to general 2D DLM problems with essentially the same complexity regardless of the metric being used. Techniques such as toggle/swap, however, are designed to exploit special properties of additive quadratic metrics. If alternative metrics were required (*e.g.*, to produce an optimal halftone under an entropy constraint) the complexity of the algorithm would increase substantially. The application of the grid algorithm to other 2D DLM problems, such as optimal near-lossless compression and entropy-constrained optimal halftoning, is an interesting area for future research.

5. REFERENCES

- [1] M. Aji, and R. J. McEliece, "The generalized distributive law," *IEEE Trans. Inform. Theory*, vol. 46, pp.325-343, March 2000.
- [2] K. M.Chugg, A. Anastasopoulos, and X. Chen *Iterative Detection.*, Kluwer Academic Publishers, 2001.
- [3] K. M. Chugg, X. Chen, A. Ortega, and C.-W. Chang, "An iterative algorithm for two-dimensional digital least metric problems with applications to digital image compression," in *Proc. IEEE ICIP'98*, Vol. 2, (Chicago, Illinois), Oct. 1998, pp. 722–726.
- [4] R. Geist, R. Reynolds, and D. Suggs, "A Markovian framework for digital halftoning," *ACM Trans. Graph.*, vol. 12, pp. 136-159, Apr. 1993.
- [5] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498-519, February 2001.
- [6] D. J. Lieberman and J. P. Allebach, "A dual interpretation for direct binary search and its implications for tone reproduction and texture quality," *IEEE Trans. Image Processing*, vol. 9, pp. 1950-1963, Nov. 2000.
- [7] D. J. Lieberman and J. P. Allebach, "Efficient model based halftoning using direct binary search," *J. Electron. Imag.*, submitted for publication
- [8] D. L. Neuhoff, T. N. Pappas, and N. Seshadri, "One-dimensional least-squares model-based halftoning," *J. Opt. Soc. Amer. A*, vol. 14, pp. 1707-1723, Aug. 1997.
- [9] T. N. Pappas and D. L. Neuhoff, "Least-squares model-based halftoning," *IEEE Trans. Image Processing*, vol. 8, pp. 1102-1116, Aug. 1999.
- [10] J. Pearl, *Probabilistic Reasoning in Intelligent Systems.*, San Mateo, CA: Morgan Kaufmann, 1988.
- [11] R. A. Ulichney, *Digital Halftoning*. Cambridge, MA: MIT Press, 1987.
- [12] P. W. Wong, "Entropy-constrained halftoning using multi-path tree coding," *IEEE Trans. Image Processing*, vol. 6, pp. 1567-1579, Nov. 1997.