Neighbourhood-consensus message passing and its potentials in image processing applications

Tijana Ružić[a], Aleksandra Pižurica[a] and Wilfried Philips[a]

[a]Ghent University, TELIN-IPI-IBBT, Sint-Pietersnieuwstraat 41, Ghent, Belgium

ABSTRACT

In this paper, a novel algorithm for inference in Markov Random Fields (MRFs) is presented. Its goal is to find approximate maximum a posteriori estimates in a simple manner by combining neighbourhood influence of iterated conditional modes (ICM) and message passing of loopy belief propagation (LBP). We call the proposed method neighbourhood-consensus message passing because a single joint message is sent from the specified neighbourhood to the central node. The message, as a function of beliefs, represents the agreement of all nodes within the neighbourhood regarding the labels of the central node. This way we are able to overcome the disadvantages of reference algorithms, ICM and LBP. On one hand, more information is propagated in comparison with ICM, while on the other hand, the huge amount of pairwise interactions is avoided in comparison with LBP by working with neighbourhoods. The idea is related to the previously developed iterated conditional expectations algorithm. Here we revisit it and redefine it in a message passing framework in a more general form. The results on three different benchmarks demonstrate that the proposed technique can perform well both for binary and multi-label MRFs without any limitations on the model definition. Furthermore, it manifests improved performance over related techniques either in terms of quality and/or speed.

Keywords: Markov random fields, iterated conditional modes, belief propagation, message passing

1. INTRODUCTION

Markov Random Fields (MRFs) have been widely used in image processing and computer vision for decades. They represent a powerful and elegant probabilistic prior model for encoding spatial information in an image.[1,2] This framework is needed in a variety of applications, such as image denoising, super-resolution, texture synthesis, image segmentation, stereo vision etc.

Using MRFs to solve these problems actually means performing Bayesian inference to estimate a set of labels of MRF nodes that meet a certain optimization criterion (e.g. maximize a posteriori probability) starting from observed data and MRF model as a spatial prior. This problem can also be viewed as energy minimization task and it is often referred to as pixel-labelling. However, it can be computationally exhaustive or even intractable due to large number of variables and loopy structure of the graph. This is why a number of approximate inference algorithms have been developed over the years. Early attempts include Monte Carlo Markov Chain samplers, such as Gibbs[2] and Metropolis sampler,[3] which are slow but find an optimal solution with high probability. Furthermore, there is iterated conditional modes (ICM)[1] algorithm that is "greedy" method that reaches only a local optimum. Although its application is limited, it gives good results for some classes of problems. More recent techniques involve graph cuts[4,5] which give very good results for different problems and even optimal results for binary MRFs in just one iteration. However, their application is limited only to a certain class of problems. Finally, there are message passing algorithms such as loopy belief propagation (LBP)[6] and tree-reweighted message passing.[7] The abilities of all these state-of-the-art inference methods broadened the application field of MRFs to more demanding tasks. An excellent overview and comparison of these methods can be found in.[8]

In this work, we are concerned with performing inference in terms of correct pixel-labelling rather than computing the exact maximum a posteriori probability (MAP), or equivalently minimum of the energy. The same point was stated in[8] where authors concluded that reaching lower energy does not necessarily mean obtaining estimates closer to the ground truth.

Further author information: (Send correspondence to Tijana Ružić)
Tijana Ružić: E-mail: tijana.ruzic@telin.ugent.be, Telephone: +32 9 264 34 16

We develop a new approach for inference in MRF model called *neighborhood-consensus message passing* (NCMP) because we keep the idea of message passing from LBP and combine it with the use of neighbourhood from ICM. Therefore, instead of computing a message between each pair of nodes, we compute one joint message that is sent from the neighbourhood to the central node. The message is the function of beliefs. The proposed approach can also be seen as a generalization of iterated conditional expectations (ICE).[9] The ICE algorithm has been developed as an extension of ICM, but despite its great potentials, it is being neglected in the recent literature. We revisit this idea here and redefine it in a message passing framework which makes it suitable for generalizations and extensions. Furthermore, we show how it outperforms ICM and works similar or better than LBP on three example applications: binary denoising, binary segmentation and super-resolution.

The paper is organized as follows. In Section 2 we set the theoretical background and explain briefly ICM and LBP as reference algorithms. Section 3 presents the proposed NCMP. Example applications and performance comparison are given in Section 4 and conclusion in Section 5.

2. BACKGROUND AND PREVIOUS WORK

MRFs are undirected graphical models suitable for representing images. They are defined as a field of random variables $\mathbf{x}$ with the Markov property $P(x_i|\mathbf{x}_{S\backslash i}) = P(x_i|\mathbf{x}_{\partial i})$, where $S\backslash i$ is the set of all variables except the variable i and ∂i is the neighbourhood of the variable i. In image processing applications, members of the set $\mathbf{x}$ are usually called hidden nodes, where label of each hidden node $x_i \in \{1, ..., L\}$. Hidden nodes are connected to observed nodes, denoted as $\mathbf{y}$, which represent given image data. Most often used neighbourhoods consist of four and eight nearest nodes. Given the Markov property, considering only interactions between pairs of neighbouring hidden nodes and under certain simplifying assumptions,[1] the joint distribution of the observation and label fields can be factorized as

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \prod_{(i,j)} \psi_{ij}(x_i, x_j) \prod_{i} \phi_i(x_i, y_i), \tag{1}$$

where Z is a normalization constant, $\psi_{ij}(x_i, x_j)$ is the compatibility function between neighbouring nodes i and j and $\phi_i(x_i, y_i)$ is the local evidence of the node i. Strictly, local evidence is conditional probability $p(y_i|x_i)$, i.e., likelihood of the measurement y_i given the label x_i. As a consequence of a Hammersley-Clifford theorem,[3] $\psi_{ij}(x_i, x_j)$ can be expressed through pairwise potential as $\psi_{ij}(x_i, x_j) \propto \exp(-V_{ij}(x_i, x_j))$, which accounts for spatial coherence of hidden nodes and is chosen, e.g., to encourage smoothness over neighbouring nodes.

2.1 Loopy belief propagation

Here we focus on the *max-product* version of the LBP algorithm that produces maximum a posteriori (MAP) estimates. For more detailed view and equivalent sum-product version, see e.g.[6] The central concepts of the algorithm are message $m_{ij}(x_j)$ from node i to node j and belief $b_i(x_i)$ of the node i defined as:

$$m_{ij}(x_j) = \alpha \max_{x_i} \{\psi_{ij}(x_i, x_j) \phi_i(x_i, y_i) \prod_{k \in \partial i:\ k \neq j} m_{ki}(x_i)\} \tag{2}$$

$$b_i(x_i) = \alpha \phi_i(x_i, y_i) \prod_{k \in \partial i} m_{ki}(x_i), \tag{3}$$

where α is a normalization constant. The messages are iteratively updated for each pair of nodes until convergence. Beliefs are computed after messages converge. The MAP estimate of a label is the one that maximizes node's belief:

$$\hat{x}_i = \arg\max_{x_i} b_i(x_i). \tag{4}$$

Hence, in this algorithm beliefs approximate a posteriori probabilities of the nodes.

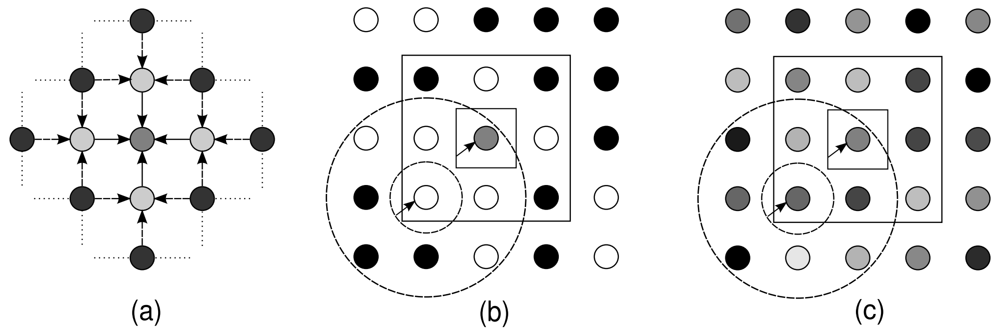

Figure 1. A graphical representation of information propagation through the graph for different algorithms. The label of the central node is to be estimated. (a) Loopy belief propagation. (b) Iterated conditional modes. (c) Neighbourhood-consensus message passing.

2.2 Iterated conditional modes

ICM is a simple, greedy inference method aiming at approximate MAP estimates. It starts from an initial estimate and then iteratively visits the nodes in some pre-defined order until convergence. While the true MAP estimate would maximize the posterior probability $p(\mathbf{x}|\mathbf{y})$, ICM in each iteration maximizes the conditional probability of x_i given the observations $\mathbf{y}$ *and* the current estimation $\hat{\mathbf{x}}_{S\setminus i}$ elsewhere:

$$\hat{x}_i = \arg\max_{x_i} p(x_i|\mathbf{y}, \hat{\mathbf{x}}_{S\setminus i}) \tag{5}$$

with

$$p(x_i|\mathbf{y}, \hat{\mathbf{x}}_{S\setminus i}) \propto \phi_i(x_i, y_i) \exp\Big(- \sum_{k\in\partial i} V_{ik}(x_i, \hat{x}_k)\Big). \tag{6}$$

In practice, this means that ICM reduces spatial context contribution to counting the number of estimated labels of each type within the neighbourhood ∂i.

3. NEIGHBOURHOOD-CONSENSUS MESSAGE PASSING

It has been previously noted[9] that ICM introduces certain loss of information due to selection of labels in each iteration. Let us have a binary MRF with two possible labels 0 and 1. If the posterior probability of a node having the label 1 was, e.g., 0.93, ICM would choose the label 1. However, it would also choose the label 1 if the posterior probability was 0.56. Owen's iterated conditional expectations (ICE) algorithm[9] deals with this problem, but has never reached the popularity of ICM and is fairly neglected in image processing community. If we denote $p(x_i|\mathbf{y}, \mathbf{x}_{S\setminus i}) = p_i(x_i)$, then the update rule for ICE, as an alternative to (6), becomes:

$$p_i(x_i) \propto \phi_i(x_i, y_i) \exp\Big(- \sum_{k\in\partial i} p_k(x_k) V_{ik}(x_i, x_k)\Big). \tag{7}$$

The assignment of labels $\hat{x}_i = \arg\max_{x_i} p_i(x_i)$ is postponed until the end of the algorithm, i.e. after the computation of $p_i(x_i)$ converges or algorithm reaches specified number of iterations.

We have reached the similar conclusion during the attempt to simplify LBP algorithm. We have noticed that LBP performs unsatisfactory for networks with huge number of nodes and short loops because messages can be stuck in these loops. In that case, they are unable to convey the necessary information globally throughout the graph. A possible reason for this behaviour might be the excessive amount of pairwise interactions. Our approach is to observe the neighbourhood as a whole entity rather than a set of nodes that individually send messages to the central node like LBP does (Fig. 1(a)). ICM also consults the whole neighbourhood, but using the labels that are fixed in each iteration (Fig. 1(b)), thus neglecting any confidence regarding their estimation.

We propose a neighbourhood-consensus message passing algorithm (NCMP) as a compromise between LBP and ICM (Fig. 1(c)). A single "message" is here formed from the whole neighbourhood (like in the case of ICM), but making use of the probabilities or "beliefs" of the neighbouring labels (similar to LBP). This message, therefore, represents the support of the whole neighbourhood for the labels of the central node and it depends on beliefs of neighbouring nodes, pairwise potentials and a spatial model being used. In general,

$$m_{\partial i \to i}(x_i) = f(\mathbf{b}_{\partial i}, V_{i,\partial i}(x_i, \mathbf{x}_{\partial i})), \tag{8}$$

where $\mathbf{b}_{\partial i} = \{b_k(x_k) : \ k \in \partial i\}$ is the set of beliefs for the labels in ∂i and $V_{i,\partial i}(x_i, \mathbf{x}_{\partial i})$ is the neighbourhood potential between the central node and its neighbourhood. For the isotropic auto-logistic MRF model,[3] we make this definition specific as

$$m_{\partial i \to i}(x_i) = \exp\Big(- \sum_{k \in \partial i} b_k(x_k) V_{ik}(x_i, x_k) \Big), \tag{9}$$

where belief according to LBP is:

$$b_k(x_k) = \alpha \phi_k(x_k, y_k) m_{\partial k \to k}(x_k). \tag{10}$$

We estimate the labels by maximizing the belief at each node like in equation (4) from Section 2.1.

By comparing (7) with (9) and (10), we can see that NCMP represents a revisit and generalization of ICE within message passing framework. Placing it into this framework opens possibilities for application in more complex MRF models, such as higher-order MRFs and anisotropic models. This is a subject of ongoing research but it is out of scope of this paper.

Note that NCMP is also an iterative algorithm, which like ICM, starts from some initial configuration. Initialization is performed by setting the incoming message for certain node to the value that favors the label of that node in the initial mask. After initialization, the algorithm runs through iterations until some stopping criterion is satisfied or until the specified number of iterations is reached.

4. EXPERIMENTS AND RESULTS

In this section, we demonstrate the potential of our algorithm on three different benchmarks. The first two are binary denoising and binary segmentation which use relatively simple MRF model. We will show that the proposed method outperforms ICM in all cases, while performing comparable or slightly better than LBP with the additional benefit of being about 10 times faster and simpler to implement. In comparison with GC, it gives comparable results but it takes more time due to its iterative nature. However, unlike GC, it is applicable to wider range of problems, which is demonstrated in the super-resolution application in the third benchmark. Implementations of LBP and GC are obtained from the websites accompanying.[8, 10]

4.1 Noise removal from a binary image

The goal is to remove noise from observed noisy image $\mathbf{y}$ whose pixel values are $y_i \in \{-1, 1\}$, $\forall i \in S$. This image is obtained by randomly flipping the sign of a certain percent of pixels in a noise-free image $\mathbf{x}$, with $x_i \in \{-1, 1\}$, $\forall i \in S$. Local evidence is given by $\phi_i(x_i, y_i) = \exp(\eta x_i y_i)$. We model $\mathbf{x}$ as an Ising MRF model where $V_{ij}(x_i, x_j) = -\gamma x_i x_j$, $\gamma > 0$, $\eta > 0$, and the neighbourhood is the second order (the first eight neighbours).

We tested the proposed method on several binary images of different sizes. A visual comparison with ICM and LBP for image from[11] of size 1259x1703 is shown on Fig. 2. The corresponding percentage of misclassified pixels is 10% for the noisy image, 6.31% for ICM, 0.42% for LBP and 0.37% for NCMP. Therefore, NCMP outperforms ICM and performs better than LBP both qualitatively and quantitatively. We also compared the result with graph cuts (GC) because it gives optimal result in one iteration. GC is indeed the fastest and with the best quantitative result (0.34% of misclassified pixels). However, it introduces an error in pixel-labelling that results in over-smoothing (e.g. the letter e in the second row) in addition to not being applicable to the

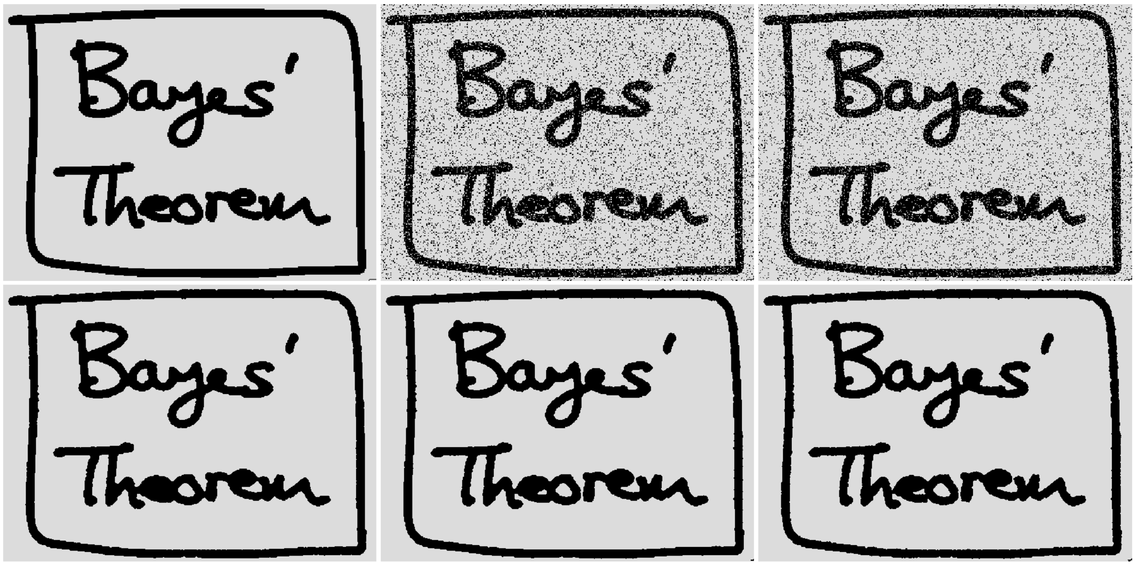

Figure 2. Top: original image, noisy image from,[11] result of ICM. Bottom: results of GC, LBP and NCMP. Parameter values: $\eta = 0.5$, $\gamma = 1.0$.

Table 1. Comparison of misclassified pixel percentage for different sizes of image from[11] ($\eta = 1$, $\gamma = 0.5$)

Percentage of the original size	ICM	GC	LBP	NCMP
100%	7.676	1.205	7.635	6.315
80%	0.359	0.071	0.233	0.144
50%	0.567	0.109	0.403	0.201

large scope of problems without modifications. Parameter values are $\eta = 0.5$ and $\gamma = 1.0$ which were the best performing ones for all the algorithms on image of this size. This is due to the huge number of nodes and loops which represents a big challenge for any inference algorithm. We also tested the algorithm on smaller input images. NCMP again performs better than ICM and LBP for certain parameter values, although the difference in performance is somewhat smaller (see Table 1).

4.2 Binary image segmentation

The second benchmark is the foreground/background segmentation algorithm from.[12] This application requires user interaction on images to be segmented (denoted as $\mathbf{y}$) to derive local evidence. User interaction files are also available with.[8] Local evidence is modelled as two separate Gaussian mixture models. The pairwise potential is a standard Potts model $V_{ij}(x_i, x_j) = -\gamma w_{ij} x_i x_j$, where $x_i, x_j \in \{0, 1\}$. This model is contrast sensitive due to $w_{i,j} = \exp(-\beta(y_i - y_j)^2)) + \lambda$, where $\beta = (2\langle(y_i - y_j)^2\rangle)^{-1}$. Values of other parameters are $\gamma = 50$ and $\lambda = 10$.

The original algorithm uses graph cuts for inference on the first order neighbourhood (the first four neighbours). Here we test its performance using the proposed method and GC, ICM and LBP as reference algorithms, in the spirit of.[8] We let all algorithms run until convergence, without specifying the initial mask. Fig. 3 shows the results for the "Flower" image, together with user interaction input. We can see that NCMP by far outperforms ICM, although it represents its simple modification. On the other hand, compared with LBP, NCMP achieves similar or slightly better accuracy of segmentation, depending on the image being processed. Additionally, NCMP is much simpler and more efficient method than LBP. Results are similar for other test images from.[8]

4.3 Super-resolution

In this super-resolution (SR) example, we wanted to show that our algorithm can work on more complicated MRF models, namely multi-label MRF with non-submodular pairwise potential. We used our SR approach

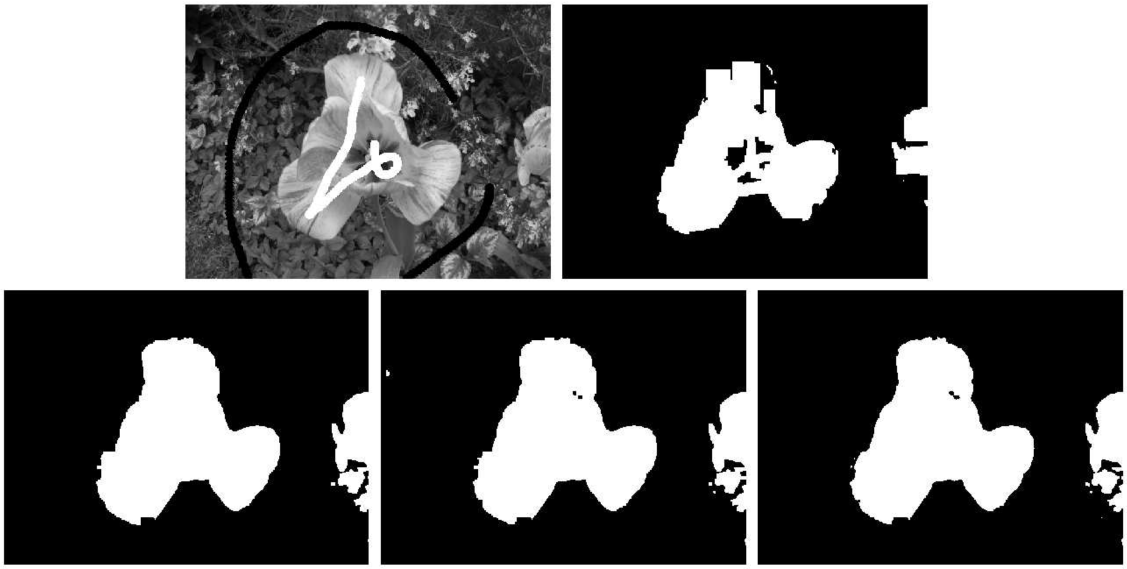

Figure 3. Top: "Flower" image with user data and the result of ICM. Bottom: result of GC, LBP and NCMP.

Figure 4. Cropped version of castle image 2x magnification. From left to right: best match result, MRF result with LBP as inference method, MRF result with NCMP as inference method.

from[13] that uses MRF model in the similar way to.[14] The idea is to model the high-resolution (HR) image as a MRF, where hidden nodes are overlapping HR patches and observed nodes are corresponding low-resolution (LR) patches. Each hidden node has a candidate set of $k > 2$ HR patches, or labels, making this a multi-label problem. Candidate set is formed by finding k nearest neighbours in some set of LR patches of each input LR patch and taking their corresponding HR patches as candidate patches. Therefore, local evidence is taken to be the matching error, i.e. sum of squared differences, between starting LR patch o_p and found k nearest neighbours y_p^n:

$$\phi_p(y_p^n, o_p) = \exp(-\|y_p^n - o_p\|^2 / 2\sigma_R^2). \tag{11}$$

and the potential is the error in the region of overlap Rov of two neighbouring HR patches:

$$V_{i,j}(x_i^n, x_j^m) = \|\mathrm{Rov}_{j,i}^n - \mathrm{Rov}_{i,j}^m\|^2. \tag{12}$$

Pairwise potential defined in this way is non-submodular[15] which means that graph cut inference method cannot be directly applied. In[15] the simplified binary form of the super-resolution from[14] was used for comparison of different inference algorithms with the proposed modification of graph cuts. Here we show that our method can perform equally well as LBP but a few times faster and without any modifications. On Fig. 4 we show the cropped version of castle image, where on the left is the result of choosing the best match at each position, i.e. when no MRF modelling is applied. In the middle and on the right are the results of LBP and the proposed method, respectively. We can see that MRF modelling brings improvement (it eliminates artefacts around edges) and that our method performs as well as LBP, while being a few times faster.

5. CONCLUSION

In this paper we have shown how ICE algorithm can be defined and generalized within the message passing framework. We have named the proposed approach neighbourhood-consensus message passing because information is propagated by sending a single joint message from the neighbourhood to the central node. Our results on three different benchmarks confirm that it can achieve better or similar results compared with reference algorithms, LBP and ICM, while preserving the simplicity of the approach. The improvement is greatly noticeable for large graphs. We believe that this particular message passing framework gives room for further generalizations and extensions.

REFERENCES

[1] Besag, J., "On the statistical analysis of dirty pictures (with discussion)," *Journal of the Royal Statistical Society. Series B* **48**, 259–302 (May 1986).

[2] Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Anal. and Machine Intel.* **6**, 721–741 (Nov 1984).

[3] Li, S. Z., [*Markov random field modeling in computer vision*], Springer-Verlag (1995).

[4] Greig, D. M., Porteous, B. T., and Seheult, A. H., "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society. Series B (Methodological)* **51**(2), 271–279 (1989).

[5] Boykov, Y., Veksler, O., and Zabih, R., "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Anal. and Machine Intel.* **23**, 1222–1239 (Nov 2001).

[6] Yedidia, J. S. and Freeman, W. T., "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Trans. on Information Theory* **47**, 736–744 (Feb 2001).

[7] Kolmogorov, V., "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. on Pattern Anal. and Machine Intel.* **28**, 1568–1583 (Oct 2006).

[8] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., and Rother, C., "A comparative study of energy minimization methods for markov random fields with smoothness-based prior," *IEEE Trans. on Pattern Anal. and Machine Intel.* **30**, 1068–1080 (Jun 2008).

[9] Owen, A., "Image segmentation via iterated conditional expectations," tech. rep., Department of Statistics, University of Chicago (1989).

[10] Tappen, M. and Freeman, W. T., "Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters," in [*Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*], 900–907 (2003).

[11] Bishop, C. M., [*Pattern recognition and machine learning*], Springer (2006).

[12] Rother, C., Kolmogorov, V., and Blake, A., "'GrabCut' - interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.* **23**, 309–314 (Aug 2004).

[13] Ružić, T., Pižurica, A., and Philips, W., "Single-image example-based super-resolution using markov random fields," in [*submitted to ACIVS2010*], (2010).

[14] Freeman, W. T., Pasztor, E. C., and Carmichael, O. T., "Learning low-level vision," *Int. J. of Computer Vision* **40**, 24–47 (Oct 2000).

[15] Rother, C., Kolmogorov, V., Lempitsky, V., and Szummer, M., "Optimizing binary mrfs via extended roof duality," in [*CVPR07*], 1–8 (2007).