Neighbourhood-Consensus Message Passing as a Framework for Generalized Iterated Conditional Expectations

Tijana Ružić, Aleksandra Pižurica and Wilfried Philips

Department for Telecommunications and Information Processing (TELIN), Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.
Tel:+32 9 264 34 12, Fax: +32 9 264 42

Abstract

In this paper we propose a novel inference method for maximum a posteriori estimation with Markov random field prior. The central idea is to integrate a kind of joint "voting" of neighbouring labels into a message passing scheme similar to loopy belief propagation (LBP). While the LBP operates with many pairwise interactions, we formulate "messages" sent from a neighbourhood as a whole. Hence the name neighbourhood-consensus message passing (NCMP). The practical algorithm is much simpler than LBP and combines the flexibility of iterated conditional modes (ICM) with some ideas of more general message passing. The proposed method is also a generalization of the iterated conditional expectations algorithm (ICE): we revisit ICE and redefine it in a message passing framework in a more general form. We also develop a simplified version of NCMP, called weighted iterated conditional modes (WICM), that is suitable for large neighbourhoods. We verify the potentials of our methods on four different benchmarks, showing the improvement in quality and/or speed over related inference techniques.

Keywords: Markov random fields, Bayesian inference, iterated conditional modes, message passing

1. Introduction

A typical problem in image processing consists of estimating some unknown image attributes from the available image data, which are incomplete or degraded. The unknown attributes can be the noise free components of the noisy image pixels, values of disparities from a stereo pair, missing pixel values, segments of the image to which each pixel belongs etc. This problem is usually referred to as *pixel-labelling*: each pixel is assigned a label representing the desired attribute. Pixel-labelling usually involves Bayesian inference like maximum a posteriori (MAP) estimation with a Markov Random Field (MRF) prior [1, 10].

In these MAP-MRF labelling problems computation is typically exhaustive or even intractable due to a large number of variables and loopy structure of the graph. Classical inference algorithms include Monte Carlo Markov Chain samplers, such as Gibbs and Metropolis sampler [10], which are slow but find an optimal solution with high probability. A popular suboptimal algorithm called iterated conditional modes (ICM) [1] is a "greedy" method that reaches only a local optimum. More recent techniques involve graph cuts (GC) [7, 4] and message passing algorithms such as loopy belief propagation (LBP) [14, 21] and tree-reweighted message passing [8]. Graph cuts give optimal solution for binary MRFs and very good result for multi-label MRFs but with the limitation of being applicable without modifications to only certain class of problems [9]. An excellent overview and comparison of these and other inference methods is in [18].

Although LBP gives state-of-the-art results in the fields of error-correcting codes [6] and computer vision (stereo matching, super-resolution [5], etc.), it has been reported to fail for graphs with huge number of nodes and many short loops [12]. We propose a novel suboptimal inference algorithm that performs well also in these cases, with the robustness of ICM and with the flexibility of a more general message passing.

We call the proposed method *neighbourhood-consensus message passing* (NCMP) because it propagates information through the graph by sending a single "consensus" message from the neighbourhood to the central node. Similarly to ICM, the message represents a unified opinion of the whole neighbourhood about the labels of the central node. Contrasting to ICM, we also take into account additional information in the form of probabilities of all neighbouring labels that form their

”voting” for the labels of the central node. Hence, the message is a function of beliefs of the neighbouring nodes representing confidence about their own labels. The proposed approach can also be considered as a generalization of iterated conditional expectations (ICE) [13] within a message passing framework. The ICE algorithm was 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, in this paper we develop another version of NCMP based on the same concept. We name this second method in the NCMP framework *weighted iterated conditional modes* (WICM) because additional information constitutes of weights added to each of the neighbouring nodes.

Experiments show that NCMP-like methods outperform ICM, while giving in many cases better or comparable results with LBP in terms of correct labelling. The improvement is greatly noticeable for large graphs with high connectivity. Moreover, our method is directly applicable to different MRF models, e.g. hierarchical or non-submodular, which is an advantage in comparison with GC. ICM has the same characteristic, but its performance is limited because it easily gets trapped in the local optimum. Despite large number of variables, our method is fast and simple to implement since it is based on local computations. Plausible result is achieved in only a few iterations. We stress that the proposed NCMP framework remains inferior to the more sophisticated LBP in more demanding applications, but achieves better performance in terms of quality and/or speed for certain simpler problems, especially when they include large graphs with high connectivity. The proposed method is definitely an interesting alternative to other low-complexity methods like ICM.

The paper is organized as follows. In Section 2 we set the theoretical background by defining MRFs. We also review briefly ICM, LBP and ICE as reference algorithms. Section 3 introduces a definition of NCMP and a novel WICM as its special case. Example applications and performance comparison are given in Section 4. Finally, we conclude the paper with Section 5.

2. Background and previous work

2.1. Markov random fields

A general image model that we consider is sketched in Fig. 1. In this representation, *observed nodes* represent given image data $\mathbf{y}$, e.g. image pixels or patches. Each observed node is connected to the corresponding *hidden node*. A set of hidden nodes is denoted as vector $\mathbf{x}$, the elements of which can take one of L values (usually referred to as labels). Thus, x_i denotes a label of node i, where $x_i \in \{1, ..., L\}$. $\mathbf{x}_A$ denotes a set of nodes with indices in the set A. The connections between the hidden nodes model their statistical dependencies and thereby prior knowledge about the image to be encoded by a Markov random field (MRF).

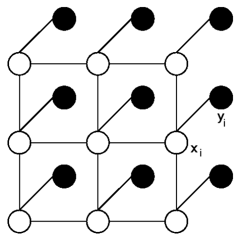

Figure 1: A square lattice of observed nodes y_i (filled circles) and hidden nodes x_i (empty circles). A pairwise MRF with the first-order neighbourhood is imposed on the hidden nodes. The edges indicate pairwise cliques.

The Markov property of a random field is formally $P(x_i|\mathbf{x}_{S\setminus i}) = P(x_i|\mathbf{x}_{\partial i})$, where $S\setminus i$ is the set of all nodes except the node i and ∂i is the neighbourhood of the node i. In words, the probability of node's label conditioned on all other labels reduces to the label's probability conditioned on its neighbours only. The distant labels have no influence on label's probability provided that its immediate neighbours are specified. Most often used neighbourhoods are first-order (four nearest nodes) and second-order neighbourhoods (eight nearest nodes). A set of nodes, which are all neighbours to one another, is called a *clique*. It can be proved that the joint distribution of a MRF $p(\mathbf{x})$ is a special case of Gibbs distribution: $p(\mathbf{x}) = 1/Z \exp\Big(-E(\mathbf{x})/T\Big)$, where T is temperature, and energy function $E(\mathbf{x}) = \sum_C V_C(\mathbf{x}_C)$. $V_C(\mathbf{x}_C)$ is called *clique potential* and is chosen in practice to favour certain local spatial dependencies, e.g. to encourage smoothness. Z is a normalization constant called partition function. In belief propagation literature it is common to consider the joint distribution between observations $\mathbf{y}$ and labels $\mathbf{x}$, which can be under certain simplifying assumptions [1] written 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 $\psi_{ij}(x_i, x_j) \propto \exp(-V_{ij}(x_i, x_j))$ denotes the statistical dependency between pairs of neighbouring hidden nodes (so called pairwise MRF) and $\phi_i(x_i, y_i)$ models relationship between observed and hidden node and is

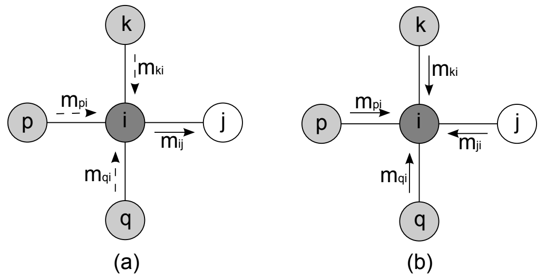

Figure 2: (a) Message update rule - the message that node i sends to node j accumulates the messages that the node i has received previously from its neighbouring nodes, other than j. (b) Belief update rule - belief of the node is calculated from *all* the incoming messages.

called local evidence. It usually stands for a conditional probability of an observed node given the value of the hidden node, $\phi_i(x_i, y_i) = p(y_i|x_i)$, and it is determined by the knowledge of the reconstruction mechanism or by learning.

2.2. Inference

The goal of Bayesian inference in MRFs is to estimate the underlying image $\mathbf{x}$ given the observed data $\mathbf{y}$ and a certain optimization criterion. Often, the criterion is to maximize the posterior probability $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x})$ with respect to $\mathbf{x}$, i.e., to compute MAP estimates. The exact inference is NP-hard, except for some special classes of problems. However, there exist a number of approximate inference algorithms that deal with this problem, as stated earlier in Section 1. Each of these has different performance in terms of quality and speed which also depends on the model being used.

2.2.1. Belief propagation

We will focus in this work on the version of the LBP algorithm called *max-product* LBP that produces maximum a posteriori (MAP) estimates. For more detailed view, see for example [21]. The central concept of the algorithm is the message defined with the message update rule as:

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

where α is a normalization constant. These messages are computed at each node and are sent to all its neighbours. Therefore, by performing local computations, the information is propagated through the graph. Message update is conducted iteratively until convergence.

The second important term in LBP is belief which is computed for each node after convergence of messages by the following equation:

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

which says that the value of node's belief depends on its local evidence and on the product of all incoming messages into the node. Beliefs actually approximate a posteriori probability of nodes. Therefore, in order to compute MAP estimates, at each node the label is chosen to maximize belief at that node:

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

In recent years various modifications of the original LBP algorithm appeared that attempt to correct some of its disadvantages. Firstly, it has been reported that LBP has poor performance for graphs with many short loops and weak evidence [12] in the sense that approximate beliefs are far from the exact ones and even that MAP estimates are incorrect. Generalized belief propagation (GBP) [22] has been proposed as solution. Here messages are exchanged between groups of nodes because those are believed to be more informative. The second modification is the tree-reweighted max-product algorithm [20] that is guaranteed to produce correct MAP estimates under certain condition. It starts from dividing the graph into a set of trees so that each edge belongs to at least one tree. Then a probability distribution over trees is chosen and each edge is given some coefficient depending on the probability that a tree contains that edge given that it contains a corresponding node. The definition of a message is then slightly modified by the introduction of this coefficient. There is also an improved version of the original algorithm, called sequential tree-reweighted message passing [8], that is guaranteed to converge. We only mention these algorithms for completeness and we do not treat them further since our approach does not build upon them.

2.2.2. Iterated conditional modes

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

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

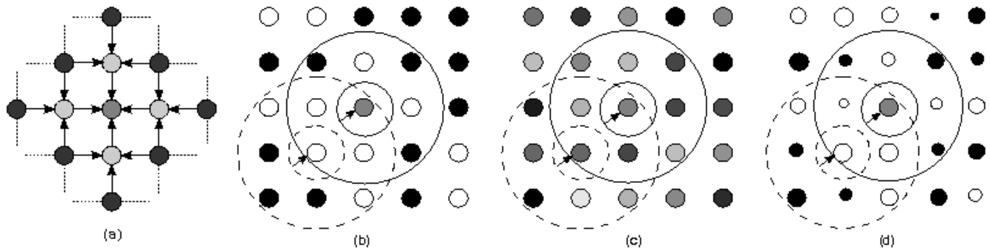

Figure 3: Graphical representation of information propagation through the binary MRF for different algorithms. The label of the central node is to be estimated. Grey node - label is to be chosen, black or white node - the label is set to one of the two values. (a) Loopy belief propagation: the messages propagate in pairwise fashion. (b) Iterated conditional modes: the decision is influenced by the estimated labels in the neighbourhood. (c) Neighbourhood-consensus message passing: the decision is influenced by the beliefs of neighbouring nodes. (d) Weighted iterated conditional modes: the decision is influenced by the estimated labels in the neighbourhood and the confidence of their estimation. Different sizes of nodes represent different weights, i.e. node's confidence about its assigned label.

Using $p(x_i|\mathbf{y}, \hat{\mathbf{x}}_{S\setminus i}) \propto \phi_i(x_i, y_i)p(x_i|\hat{\mathbf{x}}_{\partial i})$ from [1], it follows:

$$\hat{x}_i = \arg\max_{x_i} \phi_i(x_i, y_i)p(x_i|\hat{\mathbf{x}}_{\partial i}). \quad (6)$$

For pairwise MRF, which is most often used, the ICM estimate becomes:

$$\hat{x}_i = \arg\max_{x_i} \phi_i(x_i, y_i) \exp\Big(- \sum_{k\in\partial i} V_{ik}(x_i, \hat{x}_k)\Big). \quad (7)$$

Therefore, ICM reduces the spatial-context information to the number of estimated labels of each type within the neighbourhood. Once all the nodes are visited, one iteration is finished. The procedure is repeated until convergence which is guaranteed to exist and in practice is very fast. However, only local maximum is reached and results highly depend on the initial estimate.

2.2.3. Iterated conditional expectations

To overcome some of the limitations of ICM, iterated conditional expectations (ICE) algorithm was introduced [13]. Initially, ICE was developed for pairwise MRF model in image segmentation applications and later was applied to image restoration [23].

ICE transforms the discrete problem of ICM into a continuous one by assigning labels only at the end of the algorithm. In the meanwhile, approximate a posteriori probabilities of labels, which we denote as $p_i(x_i)$, are updated. The ICE update rule for pairwise MRF is then defined as

$$p_i(x_i) \propto \phi_i(x_i, y_i) \exp\Big(- \sum_{k\in\partial i} p_k(x_k)V_{ik}(x_i, x_k)\Big). \quad (8)$$

After a given stopping criterion (convergence or a predefined number of iterations), the labels are chosen so to maximize these approximate posterior probabilities: $\hat{x}_i = \arg\max_{x_i} p_i(x_i)$, i.e.

$$\hat{x}_i = \arg\max_{x_i} \phi_i(x_i, y_i) \exp\Big(-\sum_{k\in\partial i} p_k(x_k)V_{ik}(x_i, x_k)\Big). \quad (9)$$

By comparing this estimate with the equivalent one for ICM from (7), it can be seen that the product $p_k(x_k)V_{ik}(x_i, x_k)$ replaces the term $V_{ik}(x_i, \hat{x}_k)$ from the ICM estimator. This means that ICE consults all labels of a neighbouring node and makes the decision based on their posterior probability and pairwise potentials whereas ICM takes into account only the pairwise potentials derived from the fixed neighbouring labels. Note that the product $p_k(x_k)V_{ik}(x_i, x_k)$ represents matrix multiplication where $p_k(x_k)$ is a vector of posterior probabilities of all labels x_k of the node k and $V_{ik}(x_i, x_k)$ is a matrix of pairwise potentials for all possible combinations of labels x_i and x_k of two (neighbouring) nodes i and k. Despite its potential, this method is much less known than ICM in image processing community and usually neglected in recent papers.

3. Neighbourhood-consensus message passing

Our idea is to simplify LBP algorithm and make it better suited for networks with huge number of nodes and short loops. In this situation, messages are unable to convey the necessary information globally throughout the graph. The solution can be to observe a larger neighbourhood but then the speed and the practicality become an issue. 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. 3(a)). ICM also consults the whole neighbourhood, but using the labels that are fixed in each iteration (Fig. 3(b)), thus neglecting any confidence regarding their estimation. We aim at finding the compromise between these two.

To derive the new algorithm, we start from interpretations of two basic terms in LBP: message and belief. The *message* $m_{ij}(x_j)$ should express the support of the sending node i for each label of the receiving neighbouring node j. In LBP this support depends on local evidence, pairwise potential between the two nodes and incoming messages from other nodes (equation (2), Fig. 2(a)): $m_{ij}(x_j) = f(\phi_i, V_{ij}, m_{ki})$, where $\forall k \in \partial i : \ k \neq j$. This message is computed for each edge, i.e. each pair of nodes, and in each direction separately. Therefore, while computing messages, neighbourhood is not observed as one complete entity.

Our idea is to consult all neighbours of a node at once in order to make a decision about its label. ICM achieves this in a simple manner: just by counting the labels of each kind that are already fixed within the neighbourhood (Fig. 3(b)). If we look at ICM as a rather simple version of message passing, then support for the labels of a central node consists of estimated labels of neighbouring nodes. It is hidden in the term $p(x_i|\hat{\mathbf{x}}_{\partial i}) = \exp\Big(-V_{i,\partial i}(x_i, \hat{\mathbf{x}}_{\partial i})\Big)$ (equation (6)) which can be viewed as a joint message that the neighbourhood sends to the central node. This way we could depart from pairwise interactions to potentially gain more freedom in defining spatial dependencies of the nodes.

The second term in LBP is node's *belief* $b_i(x_i)$, which can be interpreted as the confidence of a node about its labels. The belief depends on the local evidence and on incoming messages into the node (equation (3), Fig. 2(b)): $b_i(x_i) = f(\phi_i, m_{ki})$, where $k \in \partial i$. For ICM, this belief is equal to one since it greedily estimates the label at each iteration being completely confident about it. This is an obvious limitation. We wish to form the messages based on the "voting" of neighbouring labels and their beliefs, in the fashion of ICE. The ICE algorithm will be a particular instance in our general approach to *neighbourhood-consensus message passing* (NCMP).

3.1. NCMP framework

The main underlying idea of our work is to propagate belief through the graph by sending a single joint message to each node from its whole neighbourhood. We define a joint message from neighbourhood ∂i to node i as a function of *neighbourhood potential* $V_{i,\partial i}(x_i, \mathbf{x}_{\partial i})$ and *neighbourhood belief* $b_{\partial i}(\mathbf{x}_{\partial i})$ (Fig. 3(c)):

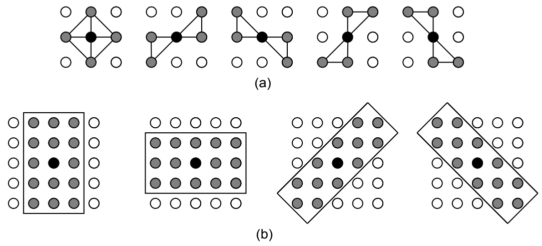

Figure 4: (a) Oriented sub-neighbourhoods from [15]. (b) Another possible set of larger larger oriented neighbourhoods.

$$m_{\partial i \to i}(x_i) = f\Big(b_{\partial i}(\mathbf{x}_{\partial i}), V_{i,\partial i}(x_i, \mathbf{x}_{\partial i})\Big). \qquad (10)$$

Note that if $\#\partial i = N$ and $x_k \in \{1, ..., L\}$, there are $\{1, ..., L^N\}$ possible label combinations for the set $\mathbf{x}_{\partial i}$. Neighbourhood potential is usually defined in MRF theory as sum of *clique potentials* $V_C(\mathbf{x}_C)$ within a given neighbourhood [10]. With this definition, the joint message in equation (10) becomes:

$$m_{\partial i \to i}(x_i) = \exp\Big(-\sum_{C \subset i \cup \partial i} b_{C\setminus i}(\mathbf{x}_{C\setminus i}) V_C(\mathbf{x}_C)\Big). \qquad (11)$$

$b_{C\setminus i}(\mathbf{x}_{C\setminus i})$ is *clique belief*, where $C\setminus i$ denotes a set of all nodes in C except the central node i. In this manner, the message $m_{\partial i \to i}(x_i)$ represents a unified opinion of the neighbourhood ∂i regarding the labelling of the central node i. Further on, in analogy to the classical belief definition from equation (3), we define node belief:

$$b_i(x_i) = \alpha \phi_i(x_i, y_i) m_{\partial i \to i}(x_i), \qquad (12)$$

where α is a normalization constant because beliefs have to sum up to one. Note that in this formulation, instead of separate messages from each neighbouring node in (3), now one joint message per neighbourhood affects the value of belief. In case of pairwise cliques, clique belief $b_{C\setminus i}(\mathbf{x}_{C\setminus i})$ from (11) reduces to node belief defined in (12). In general, for cliques containing more than two sites, we average the beliefs from (12) of the corresponding nodes:

$$b_{C\setminus i}(\mathbf{x}_{C\setminus i}) = \beta \sum_{k \in C\setminus i} b_k(x_k), \qquad (13)$$

where $\beta = \frac{1}{\#C\setminus i}$.

By comparing (11) and (12) with (8), we can see that NCMP represents a generalization of ICE in addition to being the novel message passing setting. Due to this generalization, we are not limited to pairwise interactions. Generalization gives room for improvement of the spatial prior by, for example, defining special clique potentials on redefined cliques within the neighbourhood or by using higher-order MRFs. We will illustrate advantages using oriented sub-neighbourhoods as in Fig. 4(a). These oriented sub-neighbourhoods can better model edges in different directions and retain the homogeneity of the structure.

Like in ICM, we need to start from some initial configuration. In practice, we form the initial mask by maximum likelihood estimation, $\hat{x}_i = \arg\max_{x_i} \phi_i(x_i, y_i)$, and then we initialize belief of each node by setting it to the value that favours the label of that node in the initial mask. After initialization, algorithm runs through iterations until some stopping criterion is satisfied or until the specified number of iterations is reached.

3.2. Weighted iterated conditional modes

So far we have defined NCMP algorithm by placing ICE in the context of message passing algorithms and generalizing it beyond the pairwise MRF. Now we introduce a simplified version of NCMP where we retain the discrete nature of ICM by assigning labels to nodes at each iteration while propagating additional information which is the uncertainty of that assignment. This scheme is much simpler and can be of special interest when working with large neighbourhoods. The message is now defined as an alternative to (11) as:

$$m_{\partial i \to i}(x_i) = \exp\Big(- \sum_{C \subset i \cup \partial i} b_{C \setminus i} V_C(x_i, \hat{\mathbf{x}}_{C \setminus i}) \Big) \quad (14)$$

$$\forall k \in \partial i : \hat{x}_k = \arg\max_{x_k} \phi_k(x_k, y_k) m_{\partial k \to k}(x_k). \quad (15)$$

Note that in equation (14) labels of neighbouring nodes are already set to the value that maximizes posterior probability forming that way a part of the joint neighbourhood message. The other part of the message is clique belief $b_{C \setminus i}$ that still has an interpretation of a confidence of a clique about labels of its nodes. However, that belief now has a single value rather than being computed for each possible combination of labels of the included nodes like in NCMP. This means that each node belief also has a single value computed as maximum of approximation of a posteriori probability:

$$b_k = \max_{x_k} \phi_k(x_k, y_k) m_{\partial k \to k}(x_k). \quad (16)$$

We can then define the clique belief, for example, as an average value of node beliefs that belong to the considered clique.

We call this algorithm weighted iterated conditional modes (WICM) because we add weights to clique potentials in the form of belief (see Fig. 3(d) for graphical representation). WICM is a simple modification of ICM that falls into the category of message passing. It follows the idea of NCMP while retaining the simplicity of ICM, especially when it comes to generalization beyond pairwise potentials. However, certain loss of information occurs in comparison with general NCMP due to sampling, i.e. choosing labels in each iteration, making that way a trade-off between complexity and qualitative performance.

For all the above mentioned algorithms, we have to make the following remark. There are two possible update schemes: parallel and sequential. For discrete algorithms, ICM and WICM, sequential approach updates the labels as nodes are being visited, while in parallel case all labels are updated at once at the end of each iteration. In the continuous case, i.e. LBP and NCMP, sequential approach uses messages and beliefs, respectively, calculated in the current iteration and parallel scheme uses that data from previous iteration. Both schemes have certain disadvantages: the former produces directional effects due to the scan order and the latter is prone to oscillations and therefore unable to converge [1]. We choose the parallel scheme because we are not concerned with convergence properties in our example applications.

4. Experiments and results

In this section we present a few example applications that illustrate the potentials of the proposed approach. We consider both binary and multi-label MRFs with second and first-order neighbourhoods. For comparison with GC [4, 9, 3] and LBP [19], we used the code available at http://vision.middlebury.edu/MRF/ that accompanies comparative study [18].

4.1. Noise removal from a binary image

In this example, the goal is to remove noise from an observed noisy image $\mathbf{y}$ whose pixel values are $y_i \in \{-1, 1\}$, $\forall i \in S$. We assume that the image is obtained by randomly flipping the sign of certain percent of pixels in a noise free image $\mathbf{x}$, $x_i \in \{-1, 1\}$, $\forall i \in S$. The correlation between the observed and hidden nodes is $\phi_i(x_i, y_i) = \exp(\eta x_i y_i)$, where η is a positive constant. The spatial correlation of hidden nodes that favours clustering of the labels of the same type is modelled by an Ising MRF with $\psi_{ij}(x_i, x_j) = \exp(-V_{ij}(x_i, x_j))$, where $V_{ij}(x_i, x_j) = -\gamma x_i x_j$ and γ is a positive constant. This is

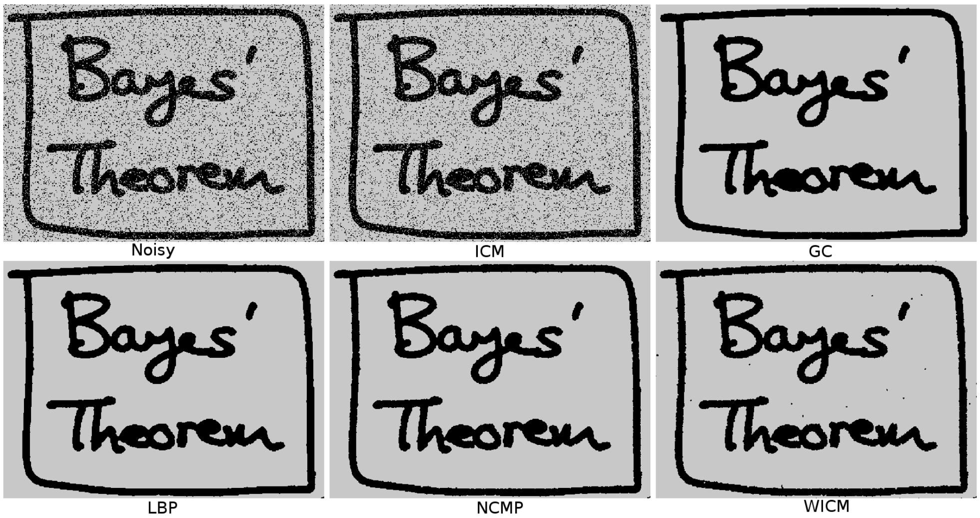

Figure 5: Top: noisy version of image from [2] of size 1259x1703 with 10% of the pixels flipped and resulting images of ICM and GC. Bottom: resulting images obtained with LBP, NCMP and WICM.

a very simple example of MRF application but is typical of more sophisticated applications and very convenient for result comparison.

The performance of the algorithm is illustrated on a binary image from [2]. The noisy version is obtained by randomly flipping 10% of the pixels in a noise free image. Denoising result depends on the value of the parameters η and γ. The only significant noise reduction is obtained for $\eta = 0.5$ and $\gamma = 1.0$, second-order neighbourhood and maximum of 500 iterations, shown in Fig. 5. It is obvious that ICM gives by far the worst result because it quickly gets trapped in the local optimum. While its performance improves on smaller images (see later Table 1), on this large image it fails completely. The proposed methods, on the contrary, work well in this case and give comparable results with the state-of-the-art methods GC and LBP. GC method gives the optimal solution for the energy function on binary MRF in only one iteration, which makes it the fastest method, but some errors in pixel-labelling are noticeable, e.g. the letter e in the second row. Finally, if we compare the two proposed methods, WICM yields slightly poorer results than NCMP because of isolated dots in the background, while still outperforming ICM.

We can also measure the quantitative performance by comparing the percentage of misclassified pixels with the original. For the above mentioned parameters, the percentage of misclassified pixels is 0.3679% for NCMP and 0.4176% for WICM compared to 6.3027% for ICM and 0.4255% for LBP. GC method yields an optimal solution of the overall best quality with only 0.3423% misclassified pixels, although visually the result of LBP and NCMP would probably be preferred (see Fig. 5). Other parameter values give significantly poorer results for all methods.

Table 1: Comparison of misclassified pixel percentage for different sizes of test image from Fig. 5 with parameters $\eta = 1$ and $\gamma = 0.5$

% of the original size	ICM	GC	LBP	NCMP	WICM
100%	7.68	1.21	7.64	6.34	6.33
80%	0.36	0.07	0.23	0.14	0.14
50%	0.57	0.11	0.4	0.2	0.2

In terms of speed, our method is slower than GC because of multiple iterations, but it is faster than LBP by an order of magnitude in addition to being simpler for implementation.

We also noted that the performance of all the methods depends on the size of the image being processed. The image on Fig. 5 is large in size, 1259x1703, so we also tested the performance on smaller images (80% and 50% of the original size). The results are summarized in Table 1. In general, for smaller images all the methods perform better and the difference in results of different methods becomes smaller.

4.2. Detection of signal of interest in wavelet domain

Another example is detection of signal of interest, i.e. meaningful edge coefficients in noisy wavelet sub-

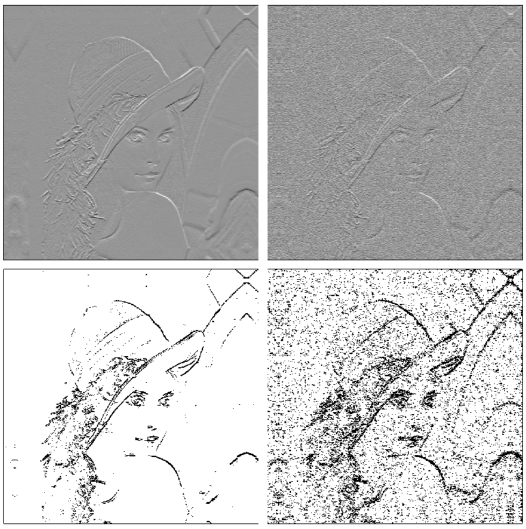

Figure 6: Top: a noise-free and the corresponding noisy wavelet sub-band (σ = 20) of Lena image. Bottom: detected edges by thresholding the two sub-bands respectively.

bands. As illustrated on Fig. 6, true edge coefficients cannot be detected by simply thresholding the noisy sub-band.

One solution is to encode prior knowledge about spatial clustering of edge coefficients using Ising MRF model [11, 15]. In this case, the labels of hidden nodes $x_i \in \{-1, 1\}$ represent absence and presence of signal of interest, respectively. Conditional likelihoods describe distributions of the magnitudes of wavelet coefficients given each label and we estimate these as described in [15]. Spatial information is given by the isotropic model with pairwise potentials $V_{ik}(x_i, x_k) = -\gamma x_i x_k$ (γ is a positive constant), that assigns a higher probability to edge continuity. We used the second-order neighbourhood. By performing inference on this model an edge map is obtained for each wavelet band that can be later used for subsequent processing in wavelet domain, e.g. for denoising.

The performance of inference algorithms LBP, ICM, NCMP and WICM on the noisy wavelet sub-band from Fig. 6 is illustrated on Fig. 7. The result was obtained for maximum of 20 iterations and $\gamma = 0.7$. Unlike in previous application from Sec. 4.1, we can see that here LBP performs poorly because it deletes most of the edges leaving the mask to be barely recognizable. On the other hand, ICM gives quite good results in this example, better than LBP. Both NCMP methods perform slightly better than ICM, yielding more consistent edges, with clearer boundaries and without interruptions. We also included the results of Metropolis sampler [10] that was originally used in [15]. The advantage of Metropolis sampler is that it estimates accurately the probabilities of each label, but in terms of the final binary mask its performance is similar to ICM. Finally, we tested GC on this model, but due to the way the model is defined, it gave no meaningful results, i.e. it recognized no signal of interest in the noisy wavelet sub-band.

Another advantage of the proposed methods and ICM in comparison with LBP, is that they can be directly applied to anisotropic models like those in Fig. 4. These models can be defined in such a way to further improve the results of edge detection. For example, in [15] the potential of sub-neighbourhood p is defined as $V_{\partial i,p}(x_i, \mathbf{x}_{\partial i,p}) = -\gamma x_i \sum_{k \in \partial i,p} x_k$ and potential of complete second-order neighbourhood is $V_{i,\partial i}(x_i, \mathbf{x}_{\partial i}) = -\gamma x_i \max_p\{\sum_{k \in \partial i,p} x_k\}$. This choice of neighbourhood potential results from the following reasoning: label $x_i = 1$ should be assigned high probability if any of the sub-neighbourhoods indicates existence of a signal of interest, while label $x_i = -1$ is given preference if none of the sub-neighbourhoods has that indication.

The benefits of using the anisotropic model were already demonstrated in [15] using Metropolis sampler. In Fig. 8 we show that the performance of the proposed NCMP method also improves largely with this anisotropic model (denoted as NCMP-A). The same is true for ICM (denoted as ICM-A). These methods give also similar results to that of the Metropolis sampler (not shown here). LBP cannot be applied directly on this hierarchical, anisotropic model and its performance is thereby inferior in this case.

4.3. Image segmentation

Another application is segmentation of a noisy image: each pixel is assigned one label that represents the segment to which the pixel belongs. In the first experiment, we used a synthetic image with artificially added white zero mean Gaussian noise of standard deviation σ. Local evidence is then the Gaussian function with mean value equal to the pixel value in the non-degraded image and standard deviation σ. The pairwise potential is determined by the discontinuity preserving Potts model $V_{i,j}(x_i, x_j) = KT(x_i \neq x_j)$, where $T()$ is one if its argument is true and zero otherwise.

The segmentation results for $\sigma = 20$, second-order neighbourhood, $K = 1$ and maximum of 50 iterations are shown on Fig. 9. ICM has the weakest performance leaving obviously misclassified areas. NCMP and WICM yield similar results to LBP. Although LBP converged in fewer iterations, it was still around five

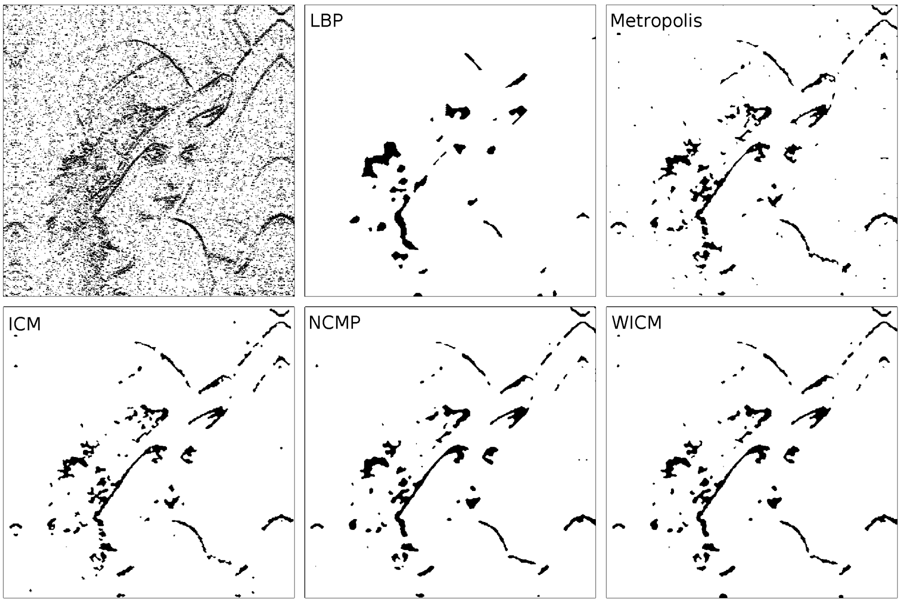

Figure 7: Masks of horizontal edges at the first scale for $\sigma = 20$. Top: initial mask and results of LBP and Metropolis sampler. Bottom: results of ICM, NCMP, and WICM.

times slower than NCMP. GC gave similar results to LBP and NCMP for the same setting.

In the second experiment, we apply the tested methods to the foreground/background segmentation from [16], where test image from Fig. 10 is used and a MRF with first-order neighbourhood. Again, our method outperforms ICM and yields a similar result to LBP in much shorter time. GC method gives the best result for this example in shortest time but note that its applicability is limited to special potential functions (see Sec. 4.4), while the proposed method is more general.

4.4. Super-resolution

For super-resolution example, we used the approach similar to [5]. The idea is to find a high-resolution (HR) patch from some candidate set for each position in the HR image so it best fits the neighbouring overlapping patches and the corresponding low-resolution (LR) patch from the input LR image. The 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. The difference with [5] is that in our case the LR and HR patches are obtained from the input image itself, rather than some external database of images. 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:

$$\phi_p(y_p^n, o_p) = \exp(-\|y_p^n - o_p\|^2 / 2\sigma_R^2). \tag{17}$$

The pairwise potential is the error in the region of overlap Rov of two neighbouring HR patches in the first-order neighbourhood:

$$V_{i,j}(x_i^n, x_j^m) = \|\mathrm{Rov}_{j,i}^n - \mathrm{Rov}_{i,j}^m\|^2. \tag{18}$$

This problem is non-submodular [17] which makes it difficult for GC. In [17] its simplified binary form was used for comparison of different inference algorithms with the proposed modification of GC. Here we keep the original problem and compare the proposed method with LBP. Fig. 11 shows the cropped version of zebra image. 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 and that our method performs equally well as LBP while being about ten times faster.

5. Conclusion and future work

In this paper, we propose a new inference method which generalizes iterated conditional expectations

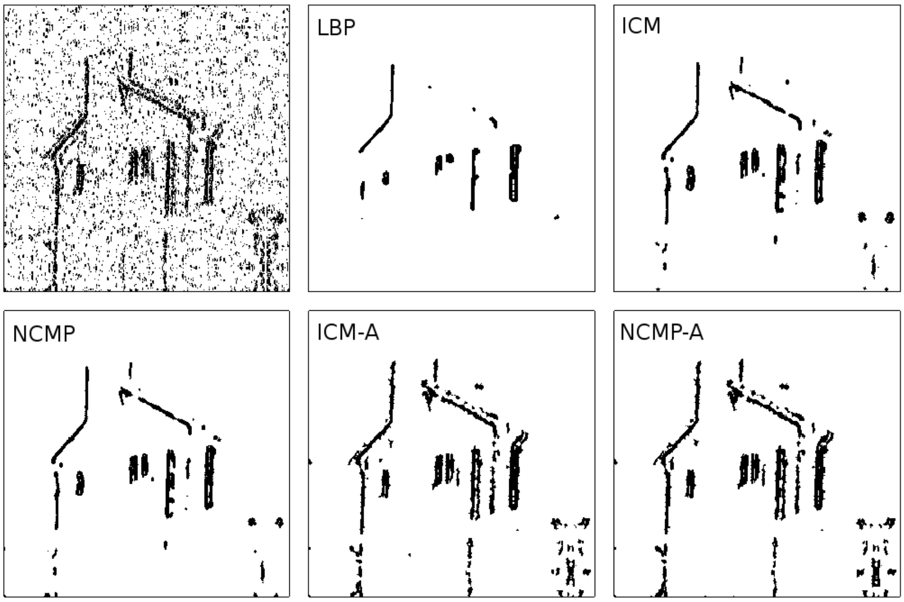

Figure 8: Masks of vertical edges at the first scale for $\sigma = 30$. Top: initial mask and results of LBP and ICM for isotropic model. Bottom: results of NCMP for isotropic model and ICM and NCMP for anisotropic model (denoted with an extension -A).

within a message passing framework. Practically, this approach is a simple modification of ICM but formulated as message passing similar to that in LBP. We call this method neighbourhood-consensus message passing (NCMP) since a joint message is sent from the specified neighbourhood to the central node which enables information to propagate through the graph. Information consists of beliefs of neighbouring nodes as confidence measure of their own labels. On the one hand, ICM is improved by adding additional information, and on the other hand, working with a whole neighbourhood is facilitated in comparison with LBP. Additionally, we develop a simplified version of NCMP, called weighted iterated conditional modes (WICM), to overcome potential difficulties while working with larger neighbourhoods. Results on different example applications show that the proposed methods outperform ICM, while giving comparable or, in some cases, even favourable results in comparison with LBP in much shorter time. Future work will focus on investigating the benefits of the proposed method for higher-order MRFs.

References

[1] Besag, J.. On the statistical analysis of dirty pictures (with discussion). Journal of the Royal Statistical Society Series B 1986;48(3):259–302.

[2] Bishop, C.M.. Pattern recognition and machine learning. Springer, 2006.

[3] Boykov, Y., Kolmogorov, V.. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell 2004;26(9):1124–1137.

[4] Boykov, Y., Veksler, O., Zabih, R.. Fast approximate energy minimization via graph cuts. IEEE Trans on Pattern Anal and Machine Intel 2001;23(11):1222–1239.

[5] Freeman, W.T., Pasztor, E.C., Carmichael, O.T.. Learning low-level vision. Int J of Computer Vision 2000;40(1):24–47.

[6] Frey, B.J., Mackay, D.J.C.. A revolution: belief propagation in graphs with cycles. In: In Neural Information Processing Systems. MIT Press; 1998. p. 479–485.

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

[8] Kolmogorov, V.. Convergent tree-reweighted message passing for energy minimization. IEEE Trans on Pattern Anal and Machine Intel 2006;28(10):1568–1583.

[9] Kolmogorov, V., Zabih, R.. What energy functions can be minimized via graph cuts? IEEE Trans on Pattern Anal and Machine Intel 2004;24(2):147–159.

[10] Li, S.Z.. Markov random field modeling in computer vision. Springer-Verlag, 1995.

[11] Malfait, M., Roose, D.. Wavelet-based image denoising using a Markov random field a priori model. IEEE Trans Image Proc 1997;6:549–565.

[12] Murphy, K.P., Weiss, Y., Jordan, M.I.. Loopy belief propagation for approximate inference: an empirical study. In: In Proceedings of Uncertainty in AI. 1999. p. 467–475.

[13] Owen, A.. Image segmentation via iterated conditional expectations. Technical Report; Department of Statistics, University of Chicago; 1989.

[14] Pearl, J.. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Fran-

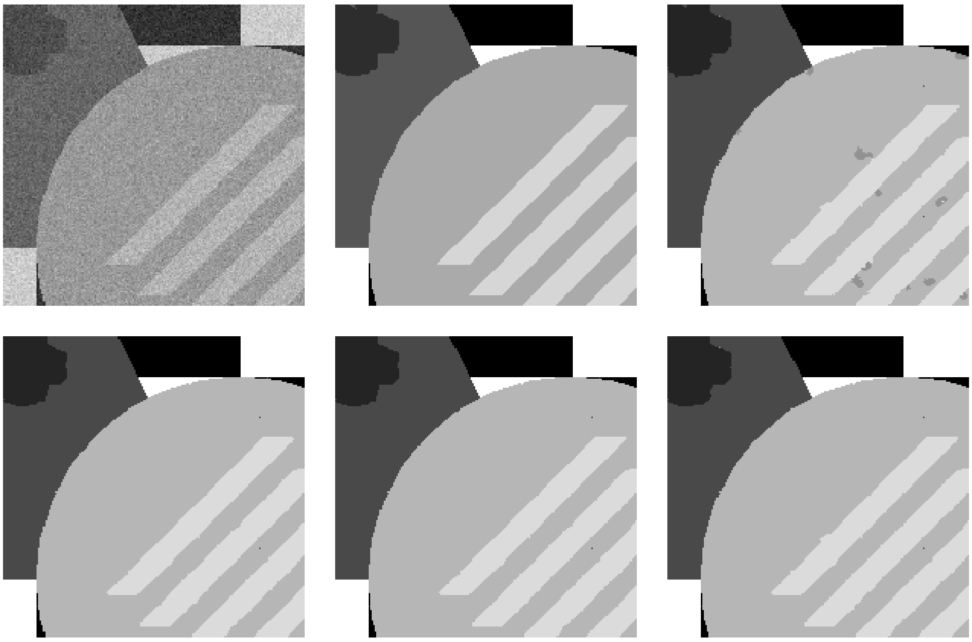

Figure 9: Cropped version of synthetic image. Top: noisy image, ground truth and result of ICM. Bottom: results of LBP, NCMP and WICM.

cisco, CA, 1988.

[15] Pižurica, A., Philips, W., Lemahieu, I., Acheroy, M.. A joint inter- and intrascale statistical model for wavelet based Bayesian image denoising. IEEE Trans Image Proc 2002;11(5):545–557.

[16] Rother, C., Kolmogorov, V., Blake, A.. 'GrabCut' - interactive foreground extraction using iterated graph cuts. ACM Trans Graph 2004;23(3):309–314.

[17] Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.. Optimizing binary mrfs via extended roof duality. In: CVPR07. 2007. p. 1–8.

[18] Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., 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 2008;30(6):1068–1080.

[19] Tappen, M., Freeman, W.T.. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In: In Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV). 2003. p. 900–907.

[20] Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.. Map estimation via agreement on trees: message-passing and linear programming. IEEE Trans on Information Theory 2005;51(11):3697–3717.

[21] Yedidia, J.S., Freeman, W.T.. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. IEEE Trans on Information Theory 2001;47(2):736–744.

[22] Yedidia, J.S., Freeman, W.T., Weiss, Y.. Generalized belief propagation. In: In NIPS 13. MIT Press; 2000. p. 689–695.

[23] Zhang, H.. Image restoration: flexible neighborhood systems and iterated conditional expections. Statistica Sinica 1993;3:117–139.

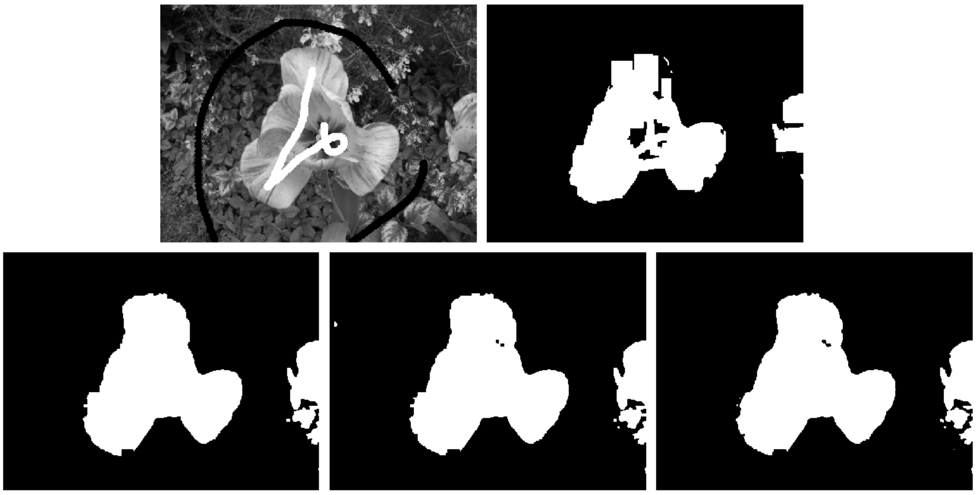

Figure 10: Binary segmentation of the flower image. Top: user data and result of ICM. Bottom: result of GC, LBP and NCMP.

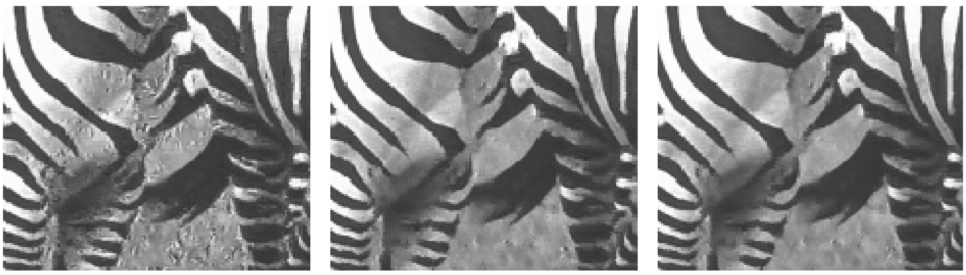

Figure 11: Cropped version of zebra image 2x magnification. From left to right: best match result, MRF result with LBP as inference method, MRF result with NCMP as inference method.