HAL open science

Inference and parameter estimation on belief networks for image segmentation

Christian Wolf, Gerald Gavin

To cite this version:

Christian Wolf, Gerald Gavin. Inference and parameter estimation on belief networks for image segmentation. Journées Francophone sur les Réseaux Bayésiens, May 2008, Lyon, France. hal-00280517

HAL Id: hal-00280517

https://hal.science/hal-00280517v1

Submitted on 19 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inference and parameter estimation on belief networks for image segmentation

Christian Wolf* — **Gérald Gavin****

Laboratoire d'informatique en images et systèmes d'information - UMR 5205

* *INSA-Lyon, Bât. J.Verne*
20, Av. Albert Einstein
69621 Villeurbanne cedex, France
christian.wolf@liris.cnrs.fr

** *Université Claude Bernard Lyon 1,*
Batiment Nautibus, 8, Boulevard Niels Bohr,
69622 Villeurbanne cedex, France
gerald.gavin@liris.cnrs.fr

ABSTRACT. *We introduce a new causal hierarchical belief network for image segmentation. Contrary to classical tree structured (or pyramidal) models, the factor graph of the network contains cycles. Each level of the hierarchical structure features the same number of sites as the base level and each site on a given level has several neighbors on the parent level. Compared to tree structured models, the (spatial) random process on the base level of the model is stationary which avoids known drawbacks, namely visual artifacts in the segmented image. We propose different parametrisations of the conditional probability distributions governing the transitions between the image levels. A parametric distribution depending on a single parameter allows the design of a fast inference algorithm on graph cuts, whereas the parameter is estimated with a least squares technique. For arbitrary distributions, we propose inference with loopy belief propagation and we introduce a new parameter estimation technique adapted to the model.*

RÉSUMÉ. *Dans ce papier, on présente un nouveau réseau bayesien hiérarchique dédié à la segmentation d'images. Contrairement aux modèles classiques (e.g. le quad arbre), le graphe associé à ce réseau bayesien contient des cycles. Chaque niveau de cette structure hiérarchique contient autant de sommets que le niveau de base (qui contient lui-même autant de sommets que l'image à segmenter contient de pixels) et chaque sommet a plusieurs parents sur le niveau supérieur. Contrairement aux structures d'arbre classiques, ce modèle possède la propriété de stationnarité ce qui signifie qu'il est invariant aux translations de l'image. Le présence de boucles dans le modèle proposé rend le problème de l'inférence exacte délicat. L'apport majeur de ce papier, outre le modèle lui-même, est la présentation d'un algorithme d'inférence exact basé sur le problème classique de graphes, à base de maximisation de flots.*

KEYWORDS: *Image segmentation, hierarchical model*

MOTS-CLÉS : *Ségmentation d'images, modèles hiérarchqiues*

1. Introduction

Image segmentation techniques aim at partitioning images into a set of non overlapping and homogeneous regions taking into account prior knowledge on the results as well as a probabilistic model of the observation (degradation) process. Belief networks, but also undirected probabilistic graphical models are widely used to incorporate spatial dependencies between the image pixels into the classification process, very often modeling the problem as Bayesian estimation.

In their seminal paper (Geman *et al.*, 1984), Geman and Geman introduced a maximum a posteriori (MAP) estimation technique for Markov random fields (MRF). An alternative to the two-dimensional MRFs are hidden Markov chains (MC) on one-dimensional traversals (Hilbert-Peano scans) of an image (Abend *et al.*, 1965) or hybrid MC/MRF techniques (Fjortoft *et al.*, 2003). The Markov chain models have been extended to belief networks with a pseudo 2D graph structure (Kuo *et al.*, 1994) and to full 2D connectivity (Levin *et al.*, 1992).

Hierarchical models introduce a scale dependent component into the classification algorithm, which allows the algorithm to better adapt itself to the image characteristics. The nodes of the graph are partitioned into different scales, where lower scale levels correspond to finer versions of the image and higher scale levels correspond to coarser versions of the image. Examples are stacks of flat MRFs (Bello, 1994), pyramidal graph structures (Kato *et al.*, 1996) and the scale causal multi-grid (Mignotte *et al.*, 2000). Bouman and Shapiro were among the first to propose a hierarchical belief network for image segmentation (Bouman *et al.*, 1994) (refined by Laferte *et al.* (Laferte *et al.*, 2000)). A quad tree models the spatial interactions between the leaf pixel sites through their interactions with neighbors in scale. The main problem of the quad tree structure is the non stationarity it induces into the random process of the leaf sites, which results in "blocky" artifacts in the segmented image.

In the same paper a second model is proposed, where each node has three parents. At first sight, the structure of the dependency graph is similar to our solution (which features four parents for each site), however, the model proposed by Bouman is a pyramidal model in that the number of nodes decreases at each level. Moreover, as an approximation the inference algorithm proposes a change of the graph after each inference step, whereas in our work the whole graph keeps its full connectivity.

The work described in this paper concentrates on the solution to the lack of shift invariance of the quad tree structured network. Our new model combines several advantages:

– Adaptation to the image characteristics with a hierarchical graph structure (similar to the quad tree)

– A stationary random process at the base level (where each site corresponds to one pixel of the input image).

– Fast inference using minimum cut/maximum flow algorithms for a subclass of transition probability distributions.

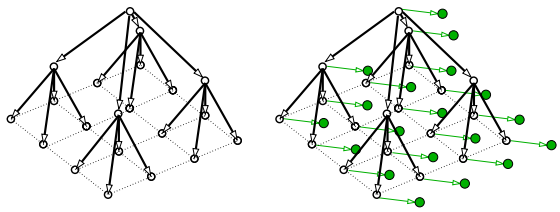


Figure 1. *The quad tree structured model with and without observed nodes.*

The paper is organized as follows: section 2 describes the quad tree structured network and section 3 extends it to the cube. Section 4 presents an inference algorithm using loopy belief propagation and section 6 presents a fast inference algorithm for a parametric class of transition probability distributions. Section 7 describes parameter estimation for the latter class of distributions and section 5 introduces an estimation technique for a nonparametric family of transition probability distributions. Section 8 discusses the computational complexity and memory requirements and section 9 experimentally valides the method. Finally, section 10 concludes.

2. Quad tree structured models

In the following we describe belief networks, thus graphical models defined on a directed acyclic graph $\mathcal{G} = \{G, E\}$, where G is a set of nodes (sites) and E is a set of edges. The edges of the graph assign, to each site s, a set of parent sites (written as s^-) and a set of children sites (written as s_-). The hierarchical nature of the graph partitions the set of nodes into levels $G^{(i)}, i \in 0..L-1$, $G^{(0)}$ being the base level corresponding to finest resolution.

Each site s is assigned a discrete random variable X_s taking values from the label set $\Lambda = \{0, \ldots, C-1\}$ where C is the number of classes. X_G, or short X denotes the field of random variables of the graph, whereas $X_{G^{(l)}}$ denotes the field of random variables at level l. The space of all possible configurations of the field X is denotated as $\Omega = \Lambda^{|G|}$.

In the case of the quad tree structured model (Bouman *et al.*, 1994; Laferte *et al.*, 2000), the graph G forms a tree structure with a single root node $r \in G^{(L-1)}$, four children nodes for each node and a single parent node for each node except the root node (see Fig. 1). Each hidden variable X_s is related to an observed variable Y_s which is conditionally independent of the other observed variables given a realization of X_s: $P(y_s|x) = P(y_s|x_s)$ and $P(y|x) = \prod_{s \in G} P(y_s|x_s)$ (see Fig. 1) . The objective is to estimate the hidden variables x given the observed variables y. In this paper we

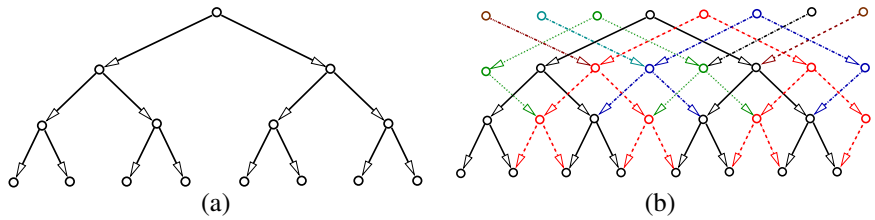


Figure 2. *A one dimensional representation of the stepwise extension of the quad tree (shown as a dyadic tree) to the cube*

consider maximum a posteriori estimation (MAP) estimation which corresponds to the mode of the posterior distribution:

$$\widehat{x} = \arg\max_{x \in \Omega} p(x|y) = \arg\max_{x \in \Omega} p(x)p(y|x) \qquad [1]$$

The absence of cycles in the dependency graph allows the application of optimization techniques similar to the ones used for the Viterbi algorithm (Jr., 1973; Viterbi, 1967) computing the optimal configuration using dynamic programming.

3. The proposed cube structured model

The main disadvantage of the Markov quad tree is the non stationarity introduced into the random process of the leaf sites $G^{(0)}$ due to the fact that, at any given level, two neighboring sites may share a common parent or not depending on their position on the grid. We therefore propose an extension shown in Fig. 2b (for easier representation the one dimensional case — a dyadic tree — is shown)

First, a second dyadic tree is added to the graph, which adds a common parent to all neighboring leaf sites which did not yet share a common parent. In the full two dimensional case, three new quad trees are added. The problem is solved for the first level, where the number of parents increased to 4 (for the full 2D model). We repeat the process for each level. New trees connect sites of the original quad tree, but also sites of the trees added at the lower levels. The final result can be seen in Fig. 2b. Note, that the final graph is not a pyramid anymore, since each level contains the same number of nodes. In general, each node has 4 parents (2 in the 1D representation) except border nodes.

The whole graph can be efficiently implemented by a cube of dimensions $N \times M \times \lceil \log_2 \max(N, M) \rceil$, $N \times M$ being the size of the image. In practice, the full

height H of the cube is not always required. The parents and children of site s with coordinates x and y on level l are given as follows:

$$s^- = \begin{cases} x+\Delta^n, y+\Delta^n \\ x+\Delta^n, y+\Delta^p \\ x+\Delta^p, y+\Delta^n \\ x+\Delta^p, y+\Delta^p \end{cases} \qquad s_- = \begin{cases} x+\Delta_n, y+\Delta_n \\ x+\Delta_n, y+\Delta_p \\ x+\Delta_p, y+\Delta_n \\ x+\Delta_p, y+\Delta_p \end{cases} \qquad [2]$$

where

$$\Delta^n = \begin{cases} -1 & \text{if } l=0 \\ -2^{l-1} & \text{else} \end{cases} \qquad \Delta^p = \begin{cases} 0 & \text{if } l=0 \\ 2^{l-1} & \text{else} \end{cases}$$
$$\Delta_n = \begin{cases} 0 & \text{if } l=1 \\ -2^{l-2} & \text{else} \end{cases} \qquad \Delta_p = \begin{cases} 1 & \text{if } l=1 \\ 2^{l-2} & \text{else} \end{cases} \qquad [3]$$

The graph as it is described in Fig. 2d (in a 1D representation) corresponds to the hidden part, i.e. the prior model $p(x)$ in the Bayesian sense. The full Markov cube including observed nodes is parametrized through three probability distributions: the discrete prior distribution of the top level labels $p(x)$, the transition probabilities $p(x_s|x_{s^-})$ and the likelihood of the observed nodes given the corresponding hidden nodes $p(y_s|x_s)$ — a probability density.

For the inference algorithm, observations at different cube levels are needed. In most cases this will require resampling the data in all levels except the finest one.

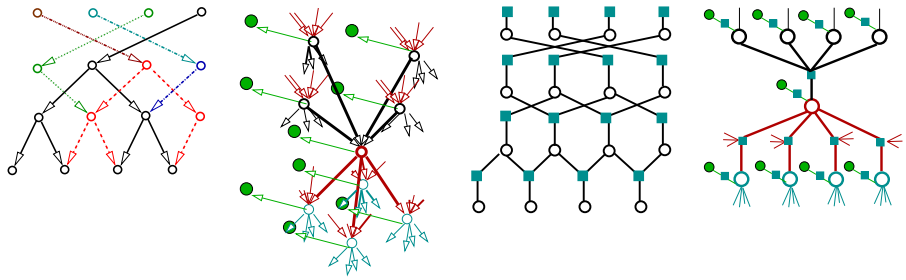
4. Inference with loopy belief propagation

Loopy belief propagation (Pearl, 1988) is an approximative inference technique for general graphs with cycles. In practice, convergence does occur for many types of graph structures. Murphy et al. present an empirical study (Murphy *et al.*, 1999) which indicates that with LBP the marginals often converge to good approximations of the posterior marginals.

Loopy belief propagation is equivalent to the sum-product (or max-product) algorithm proposed for factor graphs (Kschischang *et al.*, 2001). Any directed or undirected dependency graph can be transformed into a factor graph which contains two different types of nodes: variable nodes corresponding to the random variables of the model and factor nodes corresponding to the factors of the joint probability distribution. Figure 3 shows the 1D representation of a Markov cube without observed nodes as well as a small part of the full 2D Markov cube with observed nodes and their corresponding factor graphs.

The sum-product algorithm operates by passing messages between the nodes of the graph, each message being a function of the corresponding variable node. Due to the nature of our graph, there are two types of messages: messages from a variable node to a factor node, and the opposite:

– messages from a variable node x_s upwards to the factor node.



(a) (b) (c) (d)

Figure 3. *A cube in 1D without obervations (a) and a small part in 2D with observations (b) and the corresponding factorgraphs (c-d).*

– messages downwards to a variable node x_s coming from the factor node.

– for each child x_c of a variable node x_s, a message from x_s downwards.

– for each child x_c of a variable node x_s, a message upwards to x_s.

The message passing schedule for the Markov Cube alternates between bottom up passes and top down passes.

5. Interpretation of the hidden variables

In this section, we propose a methodology to estimate the conditional probability distributions $p(x_s|x_{s^-})$ by taking into account statistical invariance of images belonging to the same corpus. Concretely, we propose to give an interpretation of the hidden variables x_s (i.e. the variables belonging to level l>0) such that :

1) the independence model given by the structure is satisfied. Given a topological numerotation of vertices, a variable x_s should be independent of all smaller index variables given its parents x_s^-.

2) the conditional probabilities are significantly different of conditional probabilities obtained on randomly binary images.

3) the conditional probabilities are close for all images of the corpus

For simplicity reasons, in the following we describe the binary case ($C = 2$), the adaptation to multiple labels is straightforward. The Let x_s be a vertex of the Markov cube and l its level. We call U_x the set of vertices of level 0 reachable by a directed path from x. U_{x_s} is a $2^l * 2^l$ square on the image. Then, we naturally define the class of x_s as the class with the maximum number of variables U_{x_s}(in case of equality, we choose the class randomly). In order to achieve estimation, we just have to compute the frequency of label 0 (resp 1) for each parent configuration. In the corpus used in

the experiments, the 3 issues claimed above were verified. This interpretions allows several strategies for unsupervised estimation of the conditional probabilities:

– Nonparametric definition of the conditional probabilities. Given initial labels at the base level, the labels at the upper levels are computed as described above and the probabilities are estimated using histograming.

– Parametric functions are fitted to the histograms. This strategy is pursued in the next section.

6. Inference with minimum cut/maximum flow

Algorithms calculating the minimum cut/maximum flow in a graph are a powerful tool able to calculate the *exact* MAP solution on a number of binary labeling problems (D.M.Greig *et al.*, 1989; Boykov *et al.*, 2004; Boykov *et al.*, 2001; Kolmogorov *et al.*, 2004) with low order polynomial complexity. It has been shown recently, that energy functions for which the optimal solution is equivalent to the minimum cut in an appropriate graph contain only "regular" terms on binary labels (Kolmogorov *et al.*, 2004), where regular means that any projection of a term $E(x_i, x_j, x_k, \dots)$ onto any subset of two of its arguments satisfies the following condition:

$$E(0,0) + E(1,1) \leq E(0,1) + E(0,1) \qquad [4]$$

In the case of the proposed model, not all energy terms are regular, especially the terms corresponding to the logarithm of the transition probabilities $\ln p(x_s|x_{s^-})$, so the general model cannot be solved with graph cuts. However, for a large sub class with interesting properties, graph cut solutions can be found. We propose a regularizing term based on the number of parent labels which are equal to the child label:

$$P(x_s|x_{s^-}) = \frac{1}{Z}\alpha_l^{\xi(x_s, x_{s^-})} \qquad [5]$$

where α_l is a parameter depending on the level l, $\xi(x_s, x_{s^-})$ is the number of labels in x^- equal to x_s and Z is a normalization constant. The such defined transition probabilities favor homogeneous regions which corresponds to the objective of an image segmentation algorithm. We then decompose it into a sum of binary terms:

$$\ln p(x_s|x_{s^-}) = \sum_{s' \in s^-} \left[(\ln \alpha)\ \delta_{x_s, x_{s'}}\right] - Z \qquad [6]$$

where $\delta_{a,b}$ is the Kronecker delta defined as 1 if $a = b$ and 0 else. It should be noted that each binary term is regular in the sense of (Kolmogorov *et al.*, 2004). Fig. 4 shows a cut graph constructed for the dependency graph of Fig. 3b: the cut with minimum cost separating source S from sink T corresponds to the exact MAP estimate for a Markovcube with binary labels ($C = 2$). Each non terminal node is connected to one of the terminal nodes with weight $|\ln p(y_s|x_s = 1)/p(y_s|x_s = 0)|$, according to the sign inside the absolute value. The weights of top level nodes s contain an additional

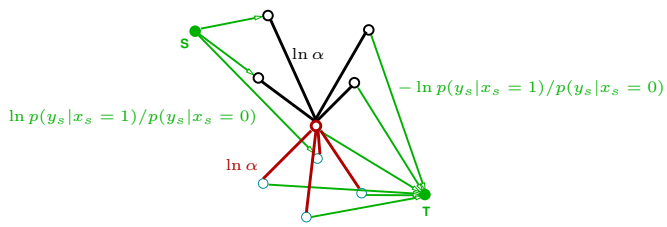


Figure 4. *The cut graph constructed for the binary problem from the dependency graph shown in Fig. 3b, including the two terminal nodes S and T. For more than 2 labels, the expansion move algorithm ressorts to a similar graph.*

term $\ln p(x_s = 1)/p(x_s = 0)$. Additionally, each non terminal node is connected to each of its parents with an undirected edge and weight $\ln \alpha$.

Minimum cut algorithms are restricted to binary labeling problems $(C = 2)$. Discontinuity preserving energy minimization with multiple labels is NP-hard (Boykov *et al.*, 2001), but the α-expansion move algorithm introduced in (Boykov *et al.*, 2001) allows to find a local minimum with garantueed maximum distance to the global minimum. It consists of iteratively applying the minimum cut algorithm to the sub problem of labeling each node of the whole graph between two labels: keeping the current label and changing the a new label α, which is changed at each iteration.

7. Parameter estimation with least squares

We chose the unsupervised technique Iterated Conditional Estimation (ICE) (Pieczynski, 2007) for parameter identification. Given supervised estimators of the parameters from a realization of the full set of variables (X, Y), an iterative procedure alternates between estimating the parameters and creating realizations of the label field based on the current parameters. The initial set of parameters can be obtained from an initial segmentation of the input image.

The prior probabilities of the top level labels β_i can be estimated using histogram techniques. Similarly, for most common observation models, maximum likelihood estimators of the sufficient statistics of the conditional distributions are readily available. In this paper, we work with a simple observation model assuming Gaussian noise, requiring as parameters means and (co)-variances for each class. Arbitrary complex likelihood functions are possible using Gaussian mixtures.

For the parameters α_l of the transition probabilities, we propose a solution based on least squares estimation similar to the works proposed by Derin et al. for the estimation of Markov random field parameters (Derin *et al.*, 1987). For each level l, we consider pairs of different site labels x_s and $x_{s'}$ $(s \in G^{(l)})$ with equal parent labels

$x_{s^-} = x_{s'^-}$. Note that the parent sites are different, whereas their labels are equal. From (5) the following relationship can be derived:

$$\frac{P(x_s|x_{s^-})}{P(x_{s'}|x_{s^-})} = \frac{\alpha_l^{\xi(x_s,x_{s^-})}}{\alpha_l^{\xi(x_{s'},x_{s^-})}} \qquad [7]$$

Expressing the conditional probabilities through absolute probabilities and taking the logarithm we get:

$$\ln \alpha_l [\, \xi(x_s, x_{s^-}) - \xi(x_{s'}, x_{s^-}) \,] = \ln \left[\frac{P(x_s, x_{s^-})}{P(x_{s'}, x_{s^-})} \right] \qquad [8]$$

The right hand side of the equation can be estimated from the label process, e.g. by histogramming, whereas the factor in the left hand side can be calculated directly. Considering a set of different label pairs, we can augment this to

$$\boldsymbol{b}^T [\ln \alpha_l] = \boldsymbol{a} \qquad [9]$$

where $\boldsymbol{b}$ is a vector where each element corresponds to the value in the left hand side of equation (8) for a given label pair and each value in the vector $\boldsymbol{a}$ corresponds to the right hand side of equation (8) for a given label pair. The solution of the over determined linear system can be found using standard least squares techniques.

8. Complexity and storage requirements

Inference complexity for loopy belief propagation (LBP) can be given as $O(I \cdot N \cdot M \cdot (H-1) \cdot C^5)$ where I is the number of iterations. H is the height of the cube and bounded by $\lceil \log_2 \max(N, M) \rceil$. Storage requires $N \cdot M \cdot (H-1) \cdot 15C$ variables. In practice, LBP in its original form is applicable for low numbers of classes (2, 3 or maximum 4), which is enough for a large number of problems. For higher numbers of classes, the classes may be quantisized and the message passing equations slightly changed.

Inference with minimum cut/maximum flow is considerable faster with a complexity bounded by $O(E * f)$, where E is the number of edges in the graph and F is the maximum flow. We use the graph cut implementation by Boykov and Kolmogorov (Boykov *et al.*, 2004) which has been optimized for typical graph structures encountered in computer vision and whose running time is nearly linear in running time in practise (Boykov *et al.*, 2001). Table 1 gives effective run times and memory requirements measured on a computer equipped with a single core Pentium-M processor running at 1.86Ghz and 1GB of RAM.

9. Experimental results

We applied the model to two common problems in document image analysis: binarization ($C = 2$) and ink bleed through, i.e. the removal of the verso side of the

Method	MB	seconds	Error rate
K-means	1	1	27.01
K-means (incl. low pass filter)	1	1	9.01
Quad tree	5	1	7.57
MRF-GC	∼20	2	6.28
Cube-LBP (4 levels, non-parametric)	103	46	6.82
Cube-LBP (4 levels, parametric)	103	46	6.91
Cube-LBP (5 levels, parametric)	150	64	6.84
Cube-GC (5 levels, parametric)	∼180	4	5.58

Table 1. *Execution times, memory requirements and segmentation performance for images of size 512×512 and (C=2)*

scanned image assuming that a verso scan is *not* available, which makes it a three class segmentation problem ($C = 3$).

In all experiments, we initialized the label field with k-means clustering after low pass filtering. The algorithms are vulnerable to numerical instabilities, we therefore resort to a widely used method calculating in the logarithmic domain.

We compared the cube model with different methods of the state of the art: flat MRF segmentation with a potts model and graph cut optimization (D.M.Greig *et al.*, 1989; Boykov *et al.*, 2001), a quad tree (Laferte *et al.*, 2000) and k-means clustering. The k-means algorithm is only method whose performance is improved when the image is low pass filtered before the segmentation.

To be able to evaluate the model quantitatively, we applied it to 30 synthetic images of size 512×512 (60 images total) and very low quality subject to multiple degradations: low pass filtering, amplification of ring shaped frequency bands causing ringing artifacts, low quality JPEG artifacts and additional Gaussian noice in various stages (with variances between σ=20 and σ=40). Table 1 shows the error rates on the different sets.

Figure 5 shows the results of the same methods applied to a real scanned document image. The images depict the restoration result, i.e. an image where the gray values of the pixels classified as "verso" have been replaced by the gray values of the nearest background pixels. As can be seen, the k-means result is noisy and the quad tree result tends to be blocky at some points. The MRF result is similar to the cube result, but tends to misclassify text pixels which are part of very fine structures.

10. Conclusion and discussion

In this paper we presented a new causal model which features the advantages of hierarchical models, i.e. scale dependent behavior and the resulting adaptivity to the image characteristics, without the main disadvantage of the quad tree model, i.e. the

Figure 5. *Restoration results on real images. From top to bottom, left to right: input image, kmeans+filtering, quad tree, MRF-GC, Cube-LBP, Cube-GC.*

lack of shift invariance. Bayesian maximum a posteriori estimation on this model has been tested on binarization and ink bleed through removal tasks for document images and compared to widely used graphical models. Segmentation quality is better or equal to the results of a MRF model, the difference depending on the scale characteristics of the input image and the nature of the degradation. We proposed two inference algorithms: loopy belief propagation and an algorithm based on graph cuts for regular transition probability distributions. In future work we will extend the cube to a conditional (discriminative) model.

11. References

Abend K., Harley T., L.N.Kanal, “ Classification of binary random Patterns”, *IEEE Transactions on Information Theory*, vol. IT-11, n° 4, p. 538-544, 1965.

Bello M., “ A combined Markov random field and wave-packet transform-based approach for image segmentation”, *IEEE transactions on image processing*, vol. 3, n° 6, p. 834-846, 1994.

Bouman C., Shapiro M., “ A Multiscale Random Field Model for Bayesian Image Segmentation”, *IEEE Transactions on Image Processing*, vol. 3, n° 2, p. 162-177, 3, 1994.

Boykov Y., Kolmogorov V., “ An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, n° 9, p. 1124-1137, 2004.

Boykov Y., Veksler O., Zabih R., " Fast Approximate Energy Minimization via Graph Cuts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, n° 11, p. 1222-1239, November, 2001.

Derin H., Elliott H., " Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, n° 1, p. 39-55, 1987.

D.M.Greig, Porteous B., Seheult A., " Exact Maximum A Posteriori Estimation for Binary Images", *Journal of the Royal Statistical Society B*, vol. 51, n° 2, p. 271-279, 1989.

Fjortoft R., Delignon Y., Pieczynski W., Sigelle M., Tupin F., " Unsupervised Classification of Radar Images Using Hidden Markov Chains and Hidden Markov Random Fields", *IEEE Transaction on Geoscience and remote sensing*, vol. 41, n° 3, p. 675-686, 2003.

Geman S., Geman D., " Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, n° 6, p. 721-741, 11, 1984.

Jr. G. F., " The Viterbi Algorithm", *Proceedings of the IEEE*, vol. 61, n° 3, p. 268-278, 1973.

Kato Z., Berthod M., Zerubia J., " A hierarchical Markov Random Field Model and Multitemperature annealing for Parallel Image Classification", *Graphical Models and Image Processing*, vol. 58, n° 1, p. 18-37, 1996.

Kolmogorov V., Zabih R., " What Energy functions can be Minimized via Graph Cuts?", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, n° 2, p. 147-159, 2004.

Kschischang F., Frey B., Loeliger H.-A., " Factor Graphs and the Sum-Product Algorithm", *IEEE transactions on Information Theory*, vol. 47, n° 2, p. 498-519, 2001.

Kuo S.-S., Agazzi O., " Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, n° 8, p. 842-848, 1994.

Laferte J.-M., Perez P., Heitz F., " Discrete Markov image modelling and inference on the quad tree", *IEEE Transactions on Image Processing*, vol. 9, n° 3, p. 390-404, 2000.

Levin E., Pieraccini R., " Dynamic Planar Warping for Optical Character Recognition", *Proceedings of the International Conference on Accoustics, Speech and Signal Processing*, vol. 3, p. 149-152, 1992.

Mignotte M., Collet C., Perez P., Bouthemy P., " Sonar Image Segmentation Using an Unsupervised Hierarchical MRF Model", *IEEE Transactions on Image Processing*, vol. 9, n° 7, p. 1216-1231, 2000.

Murphy K., Weiss Y., Jordan M. I., " Loopy belief-propagation for approximate inference: An empirical study", *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, p. 467-475, 1999.

Pearl J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufman, San Mateo, 1988.

Pieczynski W., " Convergence of the Iterative Conditional Estimation and Application to Mixture Proportion Identification", *IEEE/SP Workshop on Statistical Signal Processing*, p. 49-53, 2007.

Viterbi A., " Error bounds for convolutional codes and an asymptotically optimal decoding algorithm", *IEEE Transactions on Information Theory*, vol. IT-13, p. 260-269, 1967.