

A NEW LOCALIZED SUPERPIXEL MARKOV RANDOM FIELD FOR IMAGE SEGMENTATION

XiaoFeng Wang, Xiao-Ping Zhang

Ryerson University
Department of Electrical & Computer Engineering
350 Victoria Street, Toronto, Ontario, CANADA, M5B 2K3.
E-mail: xwang, xzhang@ee.ryerson.ca

ABSTRACT

In this paper, we present a novel localized Markov random field (MRF) method based on superpixels for region segmentation. Early vision problems could be formulated as pixel labeling using MRF. But the local interaction in MRF is limited to pixel label comparison. We propose a new localized superpixel Markov random field (SMRF) model to incorporate local data interaction in unsupervised parameter learning. The advantages of the new model include computational efficiency by using superpixel structure and its ability to integrate local knowledge in the learning process. Quantitative evaluation and visual effects show that the new model achieves not only better segmentation accuracy but also lower computational cost than the baseline pixel based model.

Index Terms— Markov random field, image segmentation, superpixel, pixel labeling

1. INTRODUCTION

Probabilistic graphical model, an interplay between probability and graph theory, provides a natural solution for uncertain and complex engineering problems [1]. Markov random field (MRF) [2] is a popular undirected graphical model used in the physics and vision field. This paper provides a new unsupervised MRF method based on superpixels to solve the region segmentation problem. Image segmentation is a fundamental problem in both image processing and computer vision. MRF based segmentation methods are very appealing for various reasons. It incorporates both prior knowledge and local spatial relationship. It has performance evaluation in a natural way. MRF methods based on pixels or regular shape neighbors are widely explored in theoretical and practical research [3]. Pixel based method is computationally complex and intractable. Recently, several superpixel based methods for supervised learning [4][5][6] are proposed. One supervised superpixel based MRF for 3D reconstruction [4] uses 4 neighbor structure feature extraction. In [5], superpixel surface layout learning does not employ MRF and uses location

as a feature. The irregular labeling scheme in [6] only considers the two neighbor pairwise structure, uses only gray level as feature and has no numerical comparison. Although superpixel concept is widely used in computer vision, it has not been applied to unsupervised learning.

In this paper a novel unsupervised superpixel based MRF (SMRF) image segmentation method is presented. There are two main contributions of this work. First, the computational efficiency is improved by applying the superpixel concept to the unsupervised image segmentation. Second, the heuristic feature difference knowledge is included in the localized parameter selection to have better segmentation accuracy. Compared to its supervised counterpart, unsupervised SMRF is faster, easier for practical use and more convenient to incorporate heuristic cues. First the image is oversegmented to contextual homogenous superpixels [7][8]. The superpixels and its neighboring superpixels form MRF graphical model. SMRF has the advantage of simplicity and accuracy over pixel based MRF. We could have reasonably good results because usually pixels in a superpixel belong to the same segment. Pixel based method usually has low accuracy because of its high dimensionality and complexity. With the new structure of SMRF, the traditional MRF model is modified to combine heuristic cues such as similarity of superpixels features for the segmentation task. The visual results and quantitative evaluation for unsupervised floor region segmentation prove the effectiveness of the method over the pixel based MRF method.

2. PROBLEM FORMULATION

Markov random field is a statistical model that is an extension of 1D Markov chain. MRF model turns out to be a very useful model in image processing areas such as image restoration, image segmentation and labeling and 3D reconstruction. The formulation of MRF model is as follows.

We denote $X = \{x_i\}_{i \in S}$ the observation data from the input image. S is a set of image sites which could be pixels or a group of adjacent pixels with regular or irregular shape.

The set of observation features at site i is x_i . The features could be any filter response of the site such as color, texture and shape. The associated labels of the image is denoted by $Y = \{y_i\}_{i \in S}$ where $y_i \in L$ is the label for site i . Here L is set of all possible labels. For example the binary image labeling problem, $L = \{0, 1\}$, 1 represents the object and 0 the background. The problem is to infer the underlying labels Y given the image features X . The probabilistic expression is $P(Y|X)$. According to Bayesian rule, MAP solution of the problem is

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)}. \quad (1)$$

Under the common assumption both the probability $P(Y)$ and the conditional probability $P(Y|X)$ are log-linear and the common factor $P(X)$ is omitted. A MRF defines a conditional probability of the following factorized form,

$$P(Y|X) = \frac{1}{Z} \exp\left\{\sum_{i \in S} f_i(x_i|y_i) + \sum_{i \in S} \sum_{j \in N_i} \beta f_{ij}(y_i, y_j)\right\} \quad (2)$$

where $f_i(\cdot)$ is the data factor of the observation features given the label of site i , $f_{ij}(\cdot)$ is the interaction factor between current site i and its neighboring site j and β is the interaction parameter. The Z is the normalize factor and $N_i \subset S$ is the neighbors of the site i . The Markov property is that site i only has interaction with its neighbor $j \in N_i$ instead of the whole set of sites S . It could also be expressed in energy terms,

$$\begin{aligned} E &= E_d + E_e = \sum_{i \in S} (E_{di} + E_{ei}) \\ &= \sum_{i \in S} f_i(x_i|y_i) + \sum_{i \in S} \sum_{j \in N_i} \beta f_{ij}(y_i, y_j). \end{aligned} \quad (3)$$

The task is to optimize the energy function to find the most probable labels. The data term E_d denotes the global features of the label and the edge term E_e is the local interaction of site labels. The E_{di} and E_{ei} are individual data term and edge term for site i . The most attractive character of the MRF is its ability to consider local and global factors at the same time.

The parameter β should be known before the model could infer the labels. In supervised learning, β is learned from the training set. In unsupervised learning, usually β is chosen experimentally without considering the features and labels. We propose a new unsupervised superpixel based model which include the feature interaction in the parameter learning.

3. THE NEW LOCALIZED SUPERPIXEL BASED MRF (SMRF) MODEL

In image processing and computer vision pixel based MRF is most commonly used. Pixel model is computational intensive and usually overfitting due to parameter estimation and optimization complexity. Since pixels are not perceptually meaningful entities of image and usually affected by noise, it is

more natural to deal with bigger contextual meaningful components. We propose a new SMRF model for unsupervised learning. Superpixel is a group of pixels that is obtained using oversegmentation methods such as normalized cut [7][8]. The benefits are computationally efficient, representationally efficient and perceptually meaningful.

The SMRF is a MRF built on superpixels instead of pixels. The underlying assumption is that the information of local superpixel is sufficient to model the distribution of whole image features. Superpixel labels at a certain location depend on just a small neighborhood of superpixels. The conditional distribution of this dependence is MRF. SMRF model assumes the superpixels are locally smooth except for discontinuities corresponding to segment boundaries and edges. Penalty is given to discontinuities. The SMRF is different from pixel based MRF in that the number of neighbors is no longer constant and the shapes of the neighbors are irregular as shown in Fig. 1. Superpixel 17 has six neighbors (13 15 22 23 25) which have different shapes.

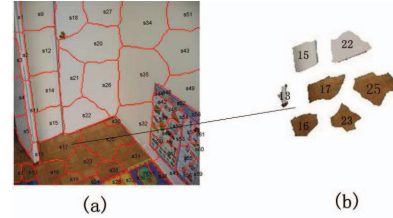


Fig. 1: New superpixel based MRF illustration.

In baseline unsupervised MRF the interaction parameter is chosen by hand. It is difficult to integrate other prior knowledge for local interaction. With the new superpixel structure we propose a new model which include adaptive parameter β considering the features and labels of the superpixel locally

$$\begin{aligned} P(Y|X) &= \frac{1}{Z} \exp\left\{\sum_{i \in S} f_i(x_i|y_i) \right. \\ &\quad \left. + \sum_{i \in S} \sum_{j \in N_i} \beta_{ij}(x_i, x_j, y_i, y_j) f_{ij}(y_i, y_j)\right\}. \end{aligned} \quad (4)$$

To simplify the problem Ising model is chosen for two class interaction function $f_{ij}(y_i, y_j) = 1 - 2\delta(y_i, y_j)$. The δ function is Kronecker delta. We assume Gaussian distribution for all features given the label, $f_i(x_i|y_i) = \sum_{k=1}^K -\frac{(x_{ik} - \bar{x}_{ik})^2}{2\delta_{x_{ik}}^2}$. Here K is the total number of features and \bar{x}_{ik} and $\delta_{x_{ik}}$ is the mean and the variance of k th feature in the feature vector given the label y_i .

We choose a new interaction parameter β based on two basic principles, the simulated annealing and the probabilistic factor of the feature difference. The parameter has the form

$$\beta_{ij}(x_i, x_j, y_i, y_j) = \frac{1}{0.9T_{t-1}(1 + \exp^{-\alpha(1-2\delta(y_i, y_j))\|x_i - x_j\|^2})}, \quad (5)$$

where T_{t-1} is annealing temperature of last iteration, t represents the current number of iterations and α is a normalize factor for feature distances which is image dependent. The factor α is set to be $(2\langle\|x_i - x_j\|^2\rangle)^{-1}$ where $\langle.\rangle$ is the average difference over the image. The common factor 0.9 is selected since the cooling speed in the range of [0.8, 0.99] is near optimal solution for annealing process [9]. By introducing new parameter function in unsupervised learning, the local interaction is weighted by feature differences. The parameter function $\beta_{ij}()$ likes a probabilistic function that reflects how the difference of two feature vectors accounts for the interaction. The chosen parameter means the higher the probability of the neighboring superpixels the higher the interaction factor and vice versa. Since the labels y_i and y_j are changed in each iteration, the new superpixel model is an adaptive algorithm with prior knowledges.

4. SMRF MODEL FOR IMAGE SEGMENTATION

To test the effectiveness of the new model, it is applied to unsupervised indoor image segmentation. First the image is oversegmented to superpixels [7]. Then SMRF model is used to infer the labels. To compare the performance between pixel based MRF (PMRF) and SMRF the EM algorithm, iterated conditional modes [10], is used for learning. Advanced energy minimization methods [11] could further improve the efficiency which is beyond the scope of our discussion. The steps are as follows,

- Step 1 - Initialization: the initial class label is assigned according to the prior knowledge of the problem. For example, in indoor image segmentation we suppose the superpixels in the lower one third part of the image is the initial floor area.
- Step 2 - Initial energy E_0 calculation: based on the initial label the total energy function of the whole image is calculated for the first time. The total energy function is the summation of the individual energy function of all superpixels. The parameters in data term could be estimated based on the labels. For example in Gaussian model the mean and the variance of each class feature are easily calculated statistically given the labels now. For the edge term the Ising model is applied and the parameter β_{ij} are calculated based on the current local feature differences shown in Equation 5.
- Step 3 - Local energy comparison: for every superpixel the local energy function for current label is compared with assigning the other class label. If it is smaller, the current label is kept otherwise the new label is assigned to the superpixel.
- Step 4 - The new total energy E_t calculation: E_t is computed with new labels and new parameters.

- Step 5 - The energy comparison: if the new total energy E_t is in small distance of the old one E_{t-1} the labeling process is finished and the algorithm converges. If the difference is still large, go to step 3 and computation continues until converges.

5. EXPERIMENTAL RESULTS

SMRF model is tested in floor area segmentation for indoor images. It is useful for indoor scene 3D reconstruction and analysis in computer vision [12]. Though it is binary labeling task, the problem is hard because in different room the floor looks very different. The ambiguity of content usually makes supervised learning unsuccessful when the room setting is unknown. A database of 53 indoor images with different layout and lighting condition is setup. They are from the MIT Labelme database [13] which is hand labeled by other researchers. Around 60 superpixels for each image is generated. The number is chosen because it is enough to have meaningful parts for the indoor images. We use the means of Lab color vector in the superpixel as color features and the edge percentage as textual feature. The initial temperature T_0 is set to be 3. The performance of the new SMRF is compared with the PMRF as shown in Fig. 2, Fig. 3 and Table 1. In Fig.2, the PMRF fails because it can not break some local minimum caused by lighting change. Since the SMRF model uses statistics of larger areas, it could proceed well. In Fig. 3, the pixel MRF catches the rug edge and some other small details, while SMRF do avoid these details. The speed of the calculation is measured by using a 2GHz HP multimedia computer with 1G RAM. The SMRF is much faster than PMRF. The reason is that we have less nodes in the SMRF model. For SMRF, the nodes are superpixels and the number is much less than the number of pixels. Precision is the number of items correctly labeled as belonging to the class divided by the total number of elements labeled as belonging to the class. Recall is defined as the number of items correctly labeled as belonging to the class divided by the total number of elements that actually belong to the class. The SMRF method has better overall results for indoor images than the PMRF in term of the accuracy and processing speed.

Table 1: SMRF and PMRF performance

Test	Precision	Recall	Speed(second)
Fig. 2(a) PMRF	57%	97%	255
Fig. 2(a) SMRF	94%	100%	77
Fig. 3(a) PMRF	68%	84%	290
Fig. 3(a) SMRF	90%	99%	78
Average PMRF	66%	84%	231
Average SMRF	78%	87%	80

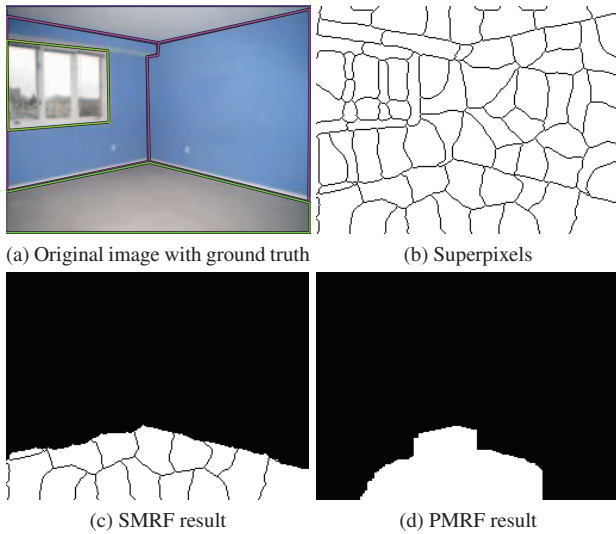


Fig. 2. SMRF and PMRF segmentation result 1.

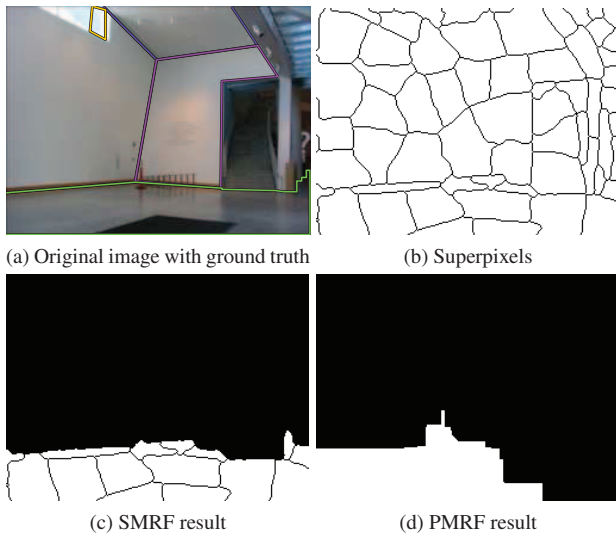


Fig. 3. SMRF and PMRF segmentation result 2.

6. CONCLUSION

In this paper, we present a novel unsupervised MRF using superpixels as nodes for image segmentation. With the new superpixel neighbor structure the local parameters are adapted in order to integrate heuristic feature difference knowledge for indoor image segmentation. The real image results prove our analysis that the new SMRF method has less computational cost and better performance for floor area segmentation than pixel based MRF. Unsupervised SMRF is a general model that could be applied to other image processing tasks other than segmentation. Future work may include using the more advanced model such as Gaussian mixture for the data term of the energy function.

7. REFERENCES

- [1] "A brief introduction to graphical models and bayesian networks," <http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html>.
- [2] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE TPAMI*, vol. 6, no. 6, pp. 721–741, 1984.
- [3] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, New York, 2001.
- [4] Ashutosh Saxena, Min Sun, and Andrew Y. Ng, "Learning 3-d scene structure from a single still image," in *ICCV workshop on 3D Representation for Recognition (3dRR-07)*, 2007.
- [5] D. Hoiem, A.A. Efros, and M. Hebert, "Recovering surface layout from an image," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 219–237, Oct. 2007.
- [6] I. Nwogu and J. J. Corso, "Labeling irregular graphs with belief propagation," in *international Workshop on Combinatorial Image Analysis*, 2008, vol. LNCS 4958.
- [7] Greg Mori, Xiaofeng Ren, Alexei A. Efros, and Jitendra Malik, "Recovering human body configurations: combining segmentation and recognition.," in *Proc. IEEE Conf. Comput. Vision and Pattern Recogn.*, 2004, vol. 2, pp. 326–333.
- [8] Xiaofeng Ren and Jitendra Malik, "Learning a classification model for segmentation," in *Proc. 9th Int'l. Conf. Computer Vision*, 2003, vol. 1, pp. 10–17.
- [9] Toshinori Munakata, *Fundamentals of the New Artificial Intelligence*, Springer-Verlag, London, 2008.
- [10] J. Besag, "On the statistical analysis of dirty images," *Journal of Royal Statistics Society*, vol. 48, pp. 259 – 302, 1986.
- [11] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, and M. Tappen C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *IEEE TPAMI*, vol. 30, no. 6, pp. 1068–1080, June 2008.
- [12] E. Delage, H. Lee, and Andrew Y. Ng, "A dynamic bayesian network model for autonomous 3d reconstruction from a single indoor image.," in *CVPR*, 2006.
- [13] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157–173, May 2008.