IMAGE BINARIZATION BASED ON CONDITIONAL RANDOM FIELDS

Yadong Mu, Bingfeng Zhou
Institute of Computer Science and Technology, Peking University, Beijing 100871, China
{muyadong, zhoubingfeng}@icst.pku.edu.cn

Abstract

In recent years, Conditional Random Fields (CRF) are proposed and proved greatly useful in natural language processing, voice recognition and computer vision. In this paper we propose a variant of CRF to solve the problem of image binarization. Unlike previous image binariztion approaches, the Patch Random Fields (PRF) proposed here could provide global optimal solutions considering both the local information from source images and pixel-wise smoothness. In this new framework, we take image patch as a kind of raw information carrier and model it with mixture of probabilistic PCA. Moreover, traditional CRF always confronts difficulties in obtaining proper parameters for the probabilistic models; this process is often time-consuming and intractable. To mitigate this problem, we train most parameters in a generative way, and then optimize the remaining parameters using a gradient descent method. The advantages of generative models and CRF are thus well combined. Experimental results demonstrate our method's effectiveness.

1. INTRODUCTION

Our work is related with three research fields: image binarization [1, 2], conditional random fields (CRF) [3-8] and patch-based methods [9].

Image binarization is an important research topic in the field of image processing. For an input image, the desired output is a black-and-white image. Image binarization is especially useful for document image analysis, scene processing, quality inspection of materials, etc. M. Sezgin and B. Sankur [1] provide an exhaustive survey of popular approaches for image binarization. They could be roughly divided into six categories: histogram shape-based methods, clustering-based methods, spatial methods, local methods, clustering-based methods and object attribute-based methods.

For the task of image binarization, information such as the local statistics could be indispensable for accurate pixel classification, which is mainly because many input images don't have uniform intensity or contrast throughout the entire image appearance. An example could be found in Fig. 1, where a global thresholding technique without local hints will fail. However, current locally adaptive thresholding approaches mostly work in a per-pixel manner, which are insufficient to handle difficult images. On the other hand, most of the global methods have difficulty in incorporating local statistics. To sum up, none of the above-mentioned approaches has utilized a framework which could combine local information and consistency within neighborhood. It is the main problem our work sets out to solve.

Conditional random fields (see [3]) provide a probabilistic framework for learning parameters and inference. Originally designed for labeling sequence data, it is later extended for voice recognition and computer vision. Various variants of CRF are proposed [5-7]. In this paper we propose a modified version of CRF, Patch Random Fields (PRF), to take advantage of the tractability of the generative models and simplify the optimization of parameters.

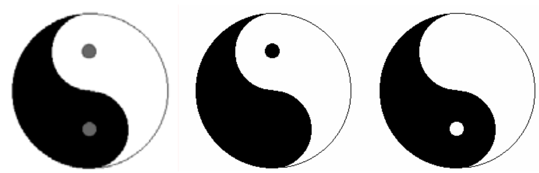

(a) (b) (c)

Fig. 1. *Illustration for the limitation of traditional global thresholding methods. The two small circles in (a) have grayscale intensity slightly less than 128, which is half of white color's. Threshold below or above 128 both fails for this special case (See (b) and (c)).*

In this paper, we choose image patch as basic information representation [9]. It is partly because patches could be efficiently calculated and carry rich information about local appearance for an image. Our observation lies in that if one knows enough about the local, then he might make decisions globally. In our work, we sample patches with predefined size from the training images as the sources of the following clustering operation. We model appearance patches with mixture of probabilistic PCA [10], which is proved efficient and effective for noise removal and dimension reduction. Under the framework of CRF, parameters are learned by maximizing pseudo likelihood.

2. PRINCIPLES OF OUR METHOD

2.1. Conditional random fields

Unlike traditional Markov random fields (MRF), conditional random fields which are proposed by Lafferty et al. [3] in 2001, directly model the posterior, rather than the joint probability of the observations and latent variables. This subtle difference relaxes the independent identical distribution (i.i.d.) assumption for distinct observation variables which tends to cause errors in many cases ([3]). Moreover, modeling the dependences among observations variables is a difficult task, sometimes even impossible. The conditional property of CRF means that it doesn't require unnecessary modeling efforts. Graphical models for MRF and CRF could be found in Fig. 2.

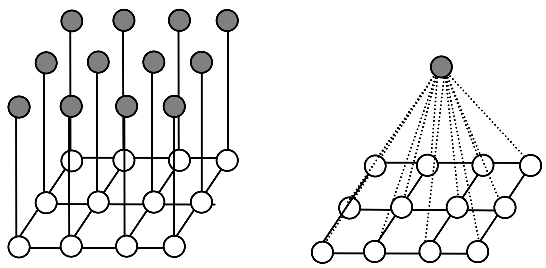


Fig. 2. *Graphical models for Markov random fields (left) and conditional random fields (right). Observation variables are colored in gray.*

We adopt the notation of [5] and [7], the definition of CRF is given as follows: For two random fields $\boldsymbol{x}$ and $\boldsymbol{y}$, where $\boldsymbol{y}$ represents latent variables and $\boldsymbol{x}$ corresponds to observations (pixel intensity, color vector, etc.), Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be a graph such that $\boldsymbol{y} = (\boldsymbol{y}_i)$, $\boldsymbol{i}$ is the node index. We said that $(\boldsymbol{x}, \boldsymbol{y})$ is a conditional random field if $\boldsymbol{y}$ obeys the Markov property:

$$p(y_i \mid x, y_{S-\{i\}}) = p(y_i \mid x, y_{N_i}) \tag{1}$$

Where $\boldsymbol{N_i}$ denotes neighbors of variable $\boldsymbol{y}_i$, and $\boldsymbol{S}\text{-}\{\boldsymbol{i}\}$ is the set of latent variables removing $\boldsymbol{y}_i$ from $\boldsymbol{S}$.

According to the Hammersley-Clifford theorem [11] and using only up to pairwise clique potentials, the posterior of $\boldsymbol{y}$ given $\boldsymbol{x}$ could be written as,

$$p(y \mid x) = \frac{1}{Z(x)} \prod_{v \in V} \varphi_V(y_v, x) \prod_{(u,v) \in E} \varphi_E(y_u, y_v, x) \tag{2}$$

Where $\mathbf{Z}(\boldsymbol{x})$ is the partition function.

2.2. Patch random fields

There are various definitions for the *association* potential and *interaction* potential (here we follow the notations in [7], the two terms correspond to the unary potential and pairwise potential in (2) respectively). In [8], J. Weinman et al. use one set of observation features for nodes and edges and transform them into feature functions. While in the work of DRF (discriminative random fields) [7] and SVRF (support vector random fields) [5], they use a Generalized Linear Model (GLM) in the form of logistic function and Support Vector Machines (SVMs) respectively.

However, the assumption of linear separability doesn't always hold, even in a very high dimensional feature space. Instead, our motivation lies in that image patches itself contain rich and raw information for many image processing and computer vision tasks, such as image binarization. Thus we directly model image patches using mixture of probabilistic PCA [10] to relax the above-mentioned assumption, rather than mapping from data space into a high dimensional feature space. Moreover, this also simplifies the optimization of parameters, which is discussed in details in Section 2.3.

2.2.1. Association potential

In our implementation, we sample all the 7*7 image patches from one or several training images, and treat them as the training dataset for a mixture of probabilistic PCA. Given the ground truth, each of these patches is cast into two predefined categories: "foreground" or "background". In this way we obtained training patches for the two mixture models: one for foreground, the other for background. The association potential could then be represented as,

$$\varphi_V(y_v, x) = \prod_{\ell} P_\ell(\bar{x}_v; \theta_\ell)^{\delta(y_{v_v}, \ell)} \tag{3}$$

Where δ is the Dirac function, and $\bar{x}_v$ is the image patch centered at node $\boldsymbol{v}$. ℓ is -1 for "background" and 1 for "foreground". θ_ℓ are parameters for PCA mixture models. P_ℓ denotes probability density function for label ℓ.

2.2.2. Interaction potential

Association potential describes the degree that observations match variable labels, while interaction potential indicates local consistency between adjacent variables. Unlike the homogeneous and isotropic Ising model used in traditional MRF, we encourage similar labels only in the sites where observations support such a consistency.

$$\varphi_E(y_u, y_v, x) = \exp(y_u y_v w \|x_u - x_v\|) \tag{4}$$

Where $\|\cdot\|$ denotes 2-norm for image pixel intensities, and $\boldsymbol{w}$ is the parameter to be optimized.

2.3. Parameter learning and inference

We uses a two-step approach to optimize parameters for potentials, which is mainly because the fact that estimating parameters for mixture models and local consistency simultaneously would be intractable.

Firstly, we ignore parameters of the Ising model and get parameters for association potential by clustering. The two clustering centers and priors are initialized by k-means and refined using iterative Expectation-Maximization algorithm.

In the second stage, parameters of interaction potential are then estimated using pseudo likelihood [7]. The posterior is approximated in a factored form:

$$P(y \mid x, w) \approx \prod_{v \in V} P(y_v \mid x, w) \tag{5}$$

$$= \prod_{v \in V} \left(\frac{1}{Z_v} \varphi_V(y_v, x) \prod_{u \in N_v} \varphi_E(y_u, y_v, x \mid w) \right) \tag{6}$$

Since it is reported [5, 7] that pseudo likelihood tends to overestimate the parameter $\boldsymbol{w}$, we thus alleviate this problem by introducing a Gaussian prior for $\boldsymbol{w}$ such that (see [7]):

$$p(w \mid \tau) = N(w : 0, \tau^2) \tag{7}$$

Where τ is a positive constant.

Given $\boldsymbol{M}$ independent training images, the interaction parameter $\boldsymbol{w}$ could then be obtained by maximizing the log-likelihood:

$$\hat{w} = \arg\max_{w} \sum_{m} \sum_{v \in V} \{ \sum_{\ell} \delta(y_v, \ell) \log P_\ell(\bar{x}_v) + \sum_{u \in N_v} y_u y_v w \| x_u - x_v \| - \log Z_v \} - \frac{1}{2\tau^2} w^2 \tag{8}$$

Where

$$Z_v = \sum_{y_v \in \{-1,1\}} \exp\{ \sum_{\ell} \delta(y_v, \ell) \log P_\ell(\bar{x}_v) + \sum_{u \in N_v} y_u y_v w \| x_u - x_v \| \} \tag{9}$$

The optimal value for $\boldsymbol{w}$ could be found by gradient descent. After we have acquired all the potential parameters, it is straightforward to assign label -1 or 1 to each pixel for a testing image using graph cuts [12].

3. EXPERIMENTAL RESULTS

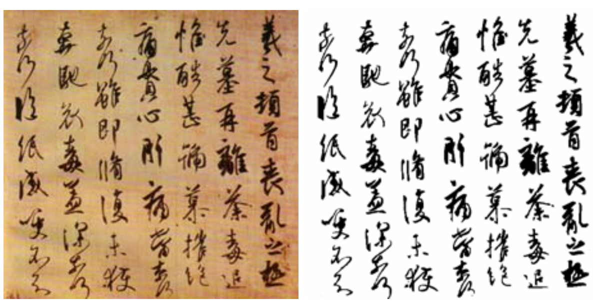
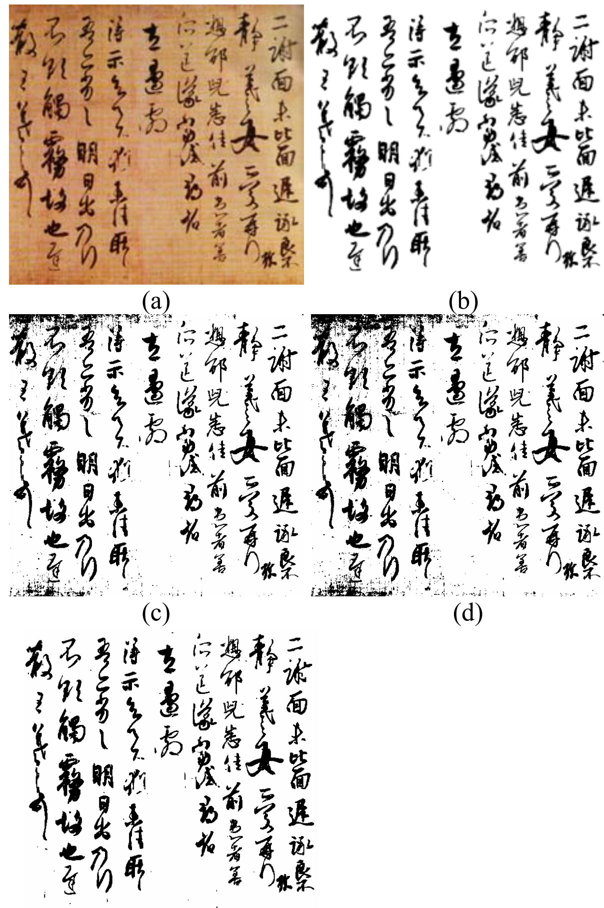


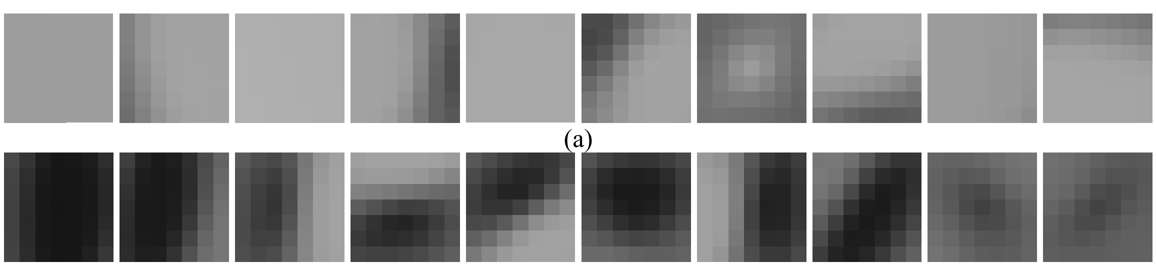
Fig. 3. An example of the training images (left) and its ground truth (right).

We collect a set of Chinese calligraphy artworks, select one or several images as training images (see Fig. 3), and keep the remaining for testing. The experimental results are given in Fig. 4.



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

Fig. 4. Experimental results for an image. (a) testing image. (b) ground truth. (c) global thresholding method with threshold set to 128. (d) Gaussian clustering-based method. (e) our method.



(a)

(b)

*Fig. 5. We represent centers of mixture of probabilistic PCA as 7*7 image patches, which are sorted in descending order according to their priors. (a) image patches for the background model. (b) image patches for the foreground model.*

We compare our method with other two methods: global thresholding method and Gaussian clustering method. As is seen in Fig. 4, our method has better visual appearance and lower classification error rate (0.0155) compared with other two(0.0244 and 0.0409 respectively). In the implementation, we use mixture models consisting of ten components. The original image patches sampled from source image have a size of 7*7, and are finally reduced to 10-dimension by PCA. Also, centers of the mixture models in the form of image patches could be found in Fig.5.

4. CONCLUSIONS

We have presented Patch Random Fields, which is a variant of CRF, and applied it for the task of image binarization. In our approach, image patches are sampled and play a central role in extracting information that is useful for our task. To reduce the difficulties of estimating parameters for these probabilistic models, we propose a two-step method, firstly learning parameters for the mixture models in a generative style, which is proven efficient. However, better methods would be explored to avoid the over-estimating effect in the optimization of the interaction potential's parameters.

REFERENCES

[1] Sezgin, M. & Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging, Society of Photo-Optical Instrumentation Engineers, 13: pp. 146-168, 2004.

[2] Wang, J.; Lee, J. & Zhang, C. Kernel GMM and its application to image binarization. Proc. of International Conference on Multimedia& Expo, 2: pp. 533-536, 2003.

[3] Lafferty, J.; McCallum, A. & Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Proc. 18th International Conf. on Machine Learning, Morgan Kaufmann, San Francisco, CA, pp. 282-289, 2001.

[4] Quattoni, A.; Collins, M. & Darrell, T. Conditional random fields for object recognition. Advances in Neural Information Processing Systems, 2004.

[5] Lee, C.; Schmidt, M.; Murtha, A.; Bistritz, A.; Sander, J. & Greiner, R. Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines. Proceedings of Workshop on Computer Vision for Biomedical Image Applications at International Conference on Computer Vision, 2005.

[6] Kumar, S. & Hebert, M. Discriminative random fields: a discriminative framework for contextual interaction in classification. Proceedings of ninth IEEE International Conference on Computer Vision, pp. 1150-1157, 2003.

[7] Kumar, S. & Hebert, M. Discriminative fields for modeling spatial dependencies in natural images. Advances in Neural Information Processing Systems, 2004.

[8] Weinman, J.; Hanson, A. & McCallum, A. Sign detection in natural images with conditional random fields. Proc. of IEEE International Workshop on Machine Learning for Signal Processing, 2004.

[9] Kannan, A.; Winn, J. & Rother, C. Clustering appearance and shape by learning jigsaws, Advances in Neural Information Processing Systems, 2006.

[10] Tipping, M. & Bishop, C. Mixtures of probabilistic principal component analyzers. Neural Computation, MIT Press, 11: pp. 443-482, 1999.

[11] Li, S. Markov random field modeling in image analysis. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2001.

[12] Boykov, Y. & Jolly, M. Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. International Conference on Computer Vision, 1: pp. 105-112, 2001.

Author Biography

Yadong Mu received the B.S. degrees from both the Computer Science Department and Philosophy Department, Peking University, Beijing, China, in 2004 and 2005 respectively. Currently, he is a Ph.D. student in the Institute of Computer Science and Technology, Peking University. His research interests include computer graphics, computational photography, computer vision and machine learning.

Bingfeng Zhou was born in 1963, Ph.D., professor, Ph.D. supervisor. His research interests include graphics simulation of robot kinematics, geometry models and CAD/CAM, color image processing, multimedia system and image special effects, digital image halftone, image based rendering and modeling, virtual reality, etc.