MAP-MRF APPROACH FOR BINARIZATION OF DEGRADED DOCUMENT IMAGE

Jung Gap Kuk, Nam Ik Cho and Kyoung Mu Lee

Seoul National University
School of Electrical Engineering
Seoul 151-744, Korea
Email : jg-kuk@ispl.snu.ac.kr, nicho@snu.ac.kr, kyoungmu@snu.ac.kr

ABSTRACT

We propose an algorithm for the binarization of document images degraded by uneven light distribution, based on the Markov Random Field modeling with Maximum *A Posteriori* probability (MAP-MRF) estimation. While the conventional algorithms use the decision based on the thresholding, the proposed algorithm makes a soft decision based on the probabilistic model. To work with the MAP-MRF framework we formulate an energy function by a likelihood model and a generalized Potts prior model. Then we construct a graph for the energy, and obtain the optimized result by using the well-known graph cut algorithm. Experimental results show that our approach is more robust to various types of images than the previous hard decision approaches.

Index Terms— binarization, graph cut, MRF, MAP

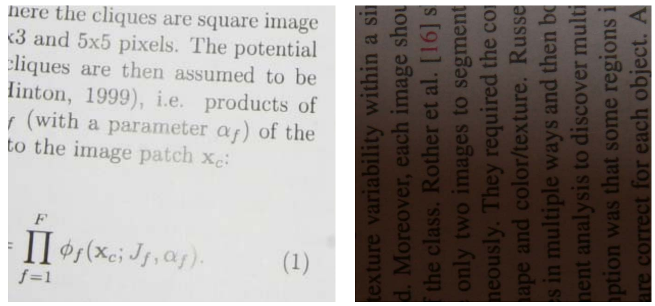
1. INTRODUCTION

Binarization of a document image is very important for its analysis, OCR, archiving with good quality, etc. Flatbed scanners are usually equipped with very efficient binarization algorithms, but they sometimes produce unreliable results near the book binding due to the change of illumination. This problem is more severe when the document is captured by a plain digital camera as shown in Fig.1, because we cannot control the light condition. There have been many adaptive binarization algorithms to tackle these problems. However, to the best of our knowledge all the algorithms are basically the thresholding algorithms that depend on the choice of thresholds :

$$f_p = \begin{cases} 1 & \text{if } \psi_p > T_p \\ 0 & \text{otherwise} \end{cases},$$

where f_p is a binary random variable which specifies an assignment of "text" or "background" to a pixel p, ψ is an information extracted from an observed image y, and T is a threshold surface.

Most of binarization algorithms for grayscale images define ψ as a pixel intensity [4, 5, 6, 2]. In this literature, how



(a) (b)

Fig. 1. Observed image with uneven light distribution

to determine a threshold surface is a key issue since the algorithm should be able to adaptively handle the local variability. The well-known Niblack algorithm adjusts a threshold level according to the statistics of pixels in a local patch [5]. The Palumbo algorithm in [6] extends the support region of statistics to the neighboring patches to have an adaptive threshold surface. Recently, a method combining these two algorithms is proposed to take advantages of both algorithms [4]. Also, a quantile linear algorithm based on the Niblack algorithm is proposed in [2].On the other hands, different information is defined in [3], which is the collection of the gradients of all edge pairs in multiscale Laplacian domain. The global threshold is determined from this information, which best discriminates character edges from the other kind of edges. In retinex based algorithm [7] , the lightness is defined as an information function and the global threshold is set experimentally.

Even though these algorithms give satisfying results for the wide range of images, they require a sophisticated and manually tuned parameters, for example, a tuning parameter α in quantile linear algorithm [2] or a ratio threshold τ in retinex based algorithm [7]. As stated in [2], the tuning procedure is very important, because *its variation influences dramatically the binarization results*. Hence, the performance is very sensitive to the choice of parameters, and the binarization result is not satisfying when the illumination change is very large. For more robust binarization, we propose a soft decision algorithm based on the MAP-MRF framework which is widely used in signal processing area. In the MAP-MRF framework we do not require a threshold surface because the algorithms work in the probabilistic world which enables soft decision. The MAP-MRF problems can be solved by the combinatorial optimization if the corresponding energy is defined. In this paper, we formulate the energy using a new likelihood model and a generalized Potts prior model, and then solve the problem by the graph cut algorithm.

This paper is organized as follows. We introduce the MAP-MRF framework for the binarization of document images in Sec.2.1. Our framework includes the estimation of light and text fields, which is presented in Sec.2.2. Experimental results are shown in Sec.3 and conclusions are given in Sec.4.

2. PROPOSED ALGORITHM

2.1. MAP-MRF framework for a binarization

The MAP estimator finds a solution which maximizes *a posteriori* probability as

$$f^* = \underset{f}{\operatorname{argmax}}\ p(f|y). \tag{1}$$

In a Bayesian framework, the conditional probability in the above equation is

$$p(f|y) \propto p(y|f)p(f),$$

where $p(y|f)$ denotes a likelihood probability and $p(f)$ denotes a prior probability. To complete the problem formulation, we need to specify $p(y|f)$ and $p(f)$. For a likelihood probability we define an observation model for a pixel p as

$$y_p = \rho_p^l(1-f_p) + \rho_p^t f_p + n. \tag{2}$$

This model states that an observation of document image is composed of a light field ρ^l and a text field ρ^t under the assumption that the observation noise n follows an independent identically distributed Gaussian. Thus a likelihood probability can be written as

$$p(y|f;\rho_p^l,\rho_p^t) = \prod_{p\in V} \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(y_p-\rho_p^l(1-f_p)-\rho_p^t f_p)^2}{2\sigma^2}}, \tag{3}$$

where V denotes a set of all pixels in the image. Note that ρ^l and ρ^t are deterministic parameters, the estimation of which will be explained in Sec.2.2. A prior probability is modeled to have a smooth labeling result. For this, we assume the random vector f to be a first order MRF which asserts that a conditional probability at a pixel depends only on the neighboring pixels. By a Hammersley-Clifford theorem a prior probability is written in an exponential family as

$$p(f) = \frac{1}{Z}\exp^{-\sum_{\{p,q\}\in E}\Psi_{p,q}(f_p,f_q)},$$

Table 1. edge weight for graph cut

edge	weight
(s,p)	$(y_p-\rho_p^l)^2$
(t,p)	$(y_p-\rho_p^t)^2$
(p,q)	$\lambda w_{p,q}\delta(f_p,f_q)$

where Z is a partition function and E denotes a set of all pairs of neighboring pixels. More Specifically, we use a *generalized* Potts model where *generalized* means *data-dependant*. It is noted that Potts model gives discontinuity preserving result. The *generalized* Potts model is usually expressed as

$$p(f) = \frac{1}{Z}\exp^{-\sum_{\{p,q\}\in E}\lambda w_{p,q}\delta(f_p,f_q)}, \tag{4}$$

where λ is for adjusting the extent of smoothness and $w_{p,q}$ is for *data-dependant*, a gradient between two pixels p and q in this paper. And $\delta(f_p,f_q)$ is a function which returns '1' when $f_p \neq f_q$ and '0' otherwise.

Incorporating (3) and (4) into (1) gives

$$\begin{aligned} f^* &= \underset{f}{\operatorname{argmax}}\ p(y|f)p(f) \\ &= \underset{f}{\operatorname{argmax}} \prod_{p\in V}\frac{1}{\sqrt{2\pi}\sigma}\exp^{-\frac{(y_p-\rho_p^l(1-f_p)-\rho_p^t f_p)^2}{2\sigma^2}} \\ &\quad \frac{1}{Z}\exp^{-\sum_{\{p,q\}\in E}\lambda w_{p,q}\delta(f_p,f_q)}, \end{aligned}$$

and taking the negative logarithm forms the minimization problem as

$$\begin{aligned} f^* &= \underset{f}{\operatorname{argmin}} \sum_{p\in V}(y_p-\rho_p^l(1-f_p)-\rho_p^t f_p)^2 \\ &\quad + \sum_{\{p,q\}\in E}\lambda w_{p,q}\delta(f_p,f_q) + const, \end{aligned} \tag{5}$$

where $\lambda \leftarrow 2\sigma^2\lambda$. In (5) we can ignore the $const$ since it doesn't depend on f. In order to minimize (5), we use the well-known graph cut algorithm [8] since it gives global optimal solution for this kind of energy. In the graph cut, a binary graph G is defined as a set of nodes $\mathcal{V} = \{V,s,t\}$ where s and t are terminal nodes and a set of edges $\mathcal{E} = \{E,E_s,E_t\}$ where $E_s = \{(s,v)|v\in V\}$ and $E_t = \{(t,v)|v\in V\}$. Edges in $\mathcal{E}$ connect the nodes in $\mathcal{V}$ under the standard 8-neighborhood system. The edge weights are computed to penalize each connection as shown in Table 1. In this construction we set a text area to the foreground ($f_p = 1$) and a light area to the background ($f_p = 0$). The conventional max-flow/min-cut algorithm on the graph $G(\mathcal{V},\mathcal{E})$ gives an optimal solution of (5).

2.2. estimation of ρ^l and ρ^t

For the completion of our model, we need to estimate the light and text fields as stated in previous section. A light field ρ^l

and a text field ρ^t should be estimated locally. To incorporate the locality of fields, we estimate the fields in a local patch $\mathcal{P}^p$ for every pixel p. We start the estimation with the classification of pixels in a patch into two groups : one for a text and the other for a light. For this, we find a set of pixel values $\mathcal{C}^p$ in $\mathcal{P}^p$. Of course, the same pixel value is not repeated in the set. If all the pixels in $\mathcal{C}^p$ belong to a light group, the plot of the sorted elements can be characterized by a line with the slope of -1. The assumption on this slope is from the fact that a light distribution is smooth in a local patch. In the case that two groups exist in a set, there might be a discontinuity which divides the set into two groups. We assume that the discontinuity is described as two lines with different slopes where one of them corresponding to a light group is still constrained to be -1. We need not consider the case that the patch covers only the text field, because the patch is usually larger than the areas of several characters. Note that although all the pixels in $\mathcal{C}^p$ belong to a light group, it can be classified as if two groups exist by a noise. However, this case can be handled in an optimization process because the prior model in eq.(4) makes it smooth.

The analysis of the set $\mathcal{C}^p$ is parameterized by $\mathbf{\Theta} = \{j, a_1(=-1), b_1, a_2, b_2\}$ where j is a boundary index dividing two groups and a and b denote the slope and the intercept for a line respectively. $\mathbf{\Theta}$ can be obtained in the MMSE (Minimum Mean Square Error) sense as

$$\mathbf{\Theta}^* = \underset{\Theta}{argmin} \quad \left(\sum_{i=0}^{j-1}(s_i-(a_1 i+b_1))^2+\sum_{i=j}^{M-1}(s_i-(a_2 i+b_2))^2\right), \tag{6}$$

where $s_i \in \mathcal{C}^p$ and $M = |\mathcal{C}^p|$. To find the solution of (6), we calculate the sum of square error e_j for each $j(=0,\ldots,M-1)$. Given $j = k$, we can easily find e_k and $\mathbf{\Theta}_k = \{j(=k), a_1^k(=-1), b_1^k, a_2^k, b_2^k\}$ by

$$\begin{aligned}
b_1^k &= \frac{\sum_{i=0}^{k-1} s_i + \sum_{i=0}^{k-1} i}{k} \\
a_2^k &= \frac{(M-k-1)\sum_{i=k}^{M-1} s_i i - \sum_{i=k}^{M-1} s_i \sum_{i=k}^{M-1} i}{(M-k-1)\sum_{i=k}^{M-1} i^2 - (\sum_{i=k}^{M-1} i)^2} \\
b_2^k &= \frac{\sum_{i=k}^{M-1} s_i - a_t^k \sum_{i=k}^{M-1} i}{M-k-1} \\
e_k &= \sum_{i=0}^{k-1}(s_i - (a_l^k i + b_l^k))^2 + \sum_{i=k}^{M-1}(s_i - (a_t^k i + b_t^k))^2.
\end{aligned}$$

Finally $\mathbf{\Theta}^*$ is obtained by

$$\mathbf{\Theta}^* = \mathbf{\Theta}_k, e_i \geq e_k \text{ for all } i.$$

Once we find the $\mathbf{\Theta}^*$ we can complete the estimation of ρ^l and ρ^t as

case 1 : $a_1 = -1$ and $a_2 = -1$
$\quad \rho_p^l = \text{mean}(\mathcal{P}_i^p, \mathcal{P}_i^p \in \mathcal{P}^p)$
$\quad \rho_p^t = y_{min}$
case 2 : otherwise
$\quad \rho_p^l = \text{mean}(\mathcal{P}_i^p, \mathcal{P}_i^p \in \mathcal{P}^p \text{ and } \mathcal{P}_i^p > s_j)$
$\quad \rho_p^t = \text{mean}(\mathcal{P}_i^p, \mathcal{P}_i^p \in \mathcal{P}^p \text{ and } \mathcal{P}_i^p \leq s_j)$.

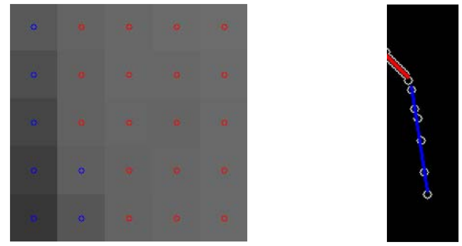

(a) local patch (b) estimated lines

Fig. 2. Estimation result when assumption succeeds

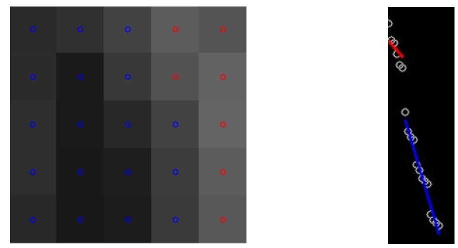
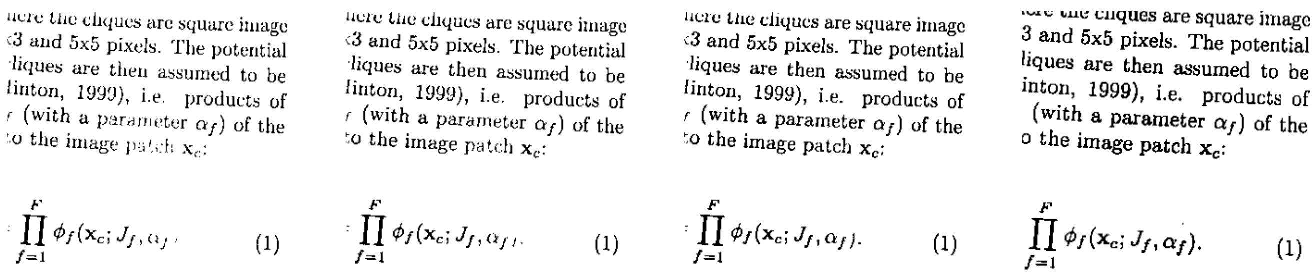

(a) local patch (b) estimated lines

Fig. 3. Estimation result when assumption fails

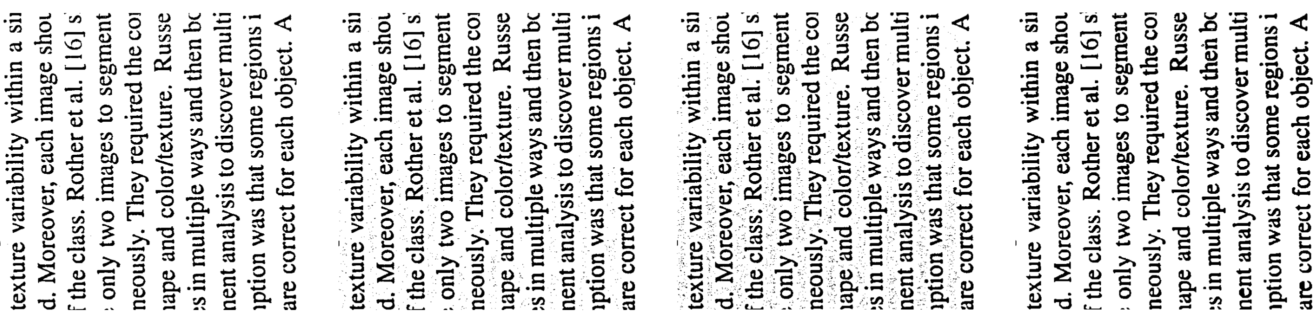
The *case 1* denotes the case when the whole pixels belong to the light group. In this case ρ_p^l is estimated by a mean of pixels in a patch and ρ_p^t is estimated by a global minimum y_{min} of an observed image. And the *case 2* denotes that there exist two groups in a patch and each field is estimated by the mean of corresponding pixels. In Fig.2, the performance of the proposed estimation method is shown, where (a) is a local patch and (b) is the estimation result. As can be seen in Fig.2 (b), a red line with the slope of -1 for a light group is well fitted to samples (white circle) and blue line with different slope for a text group is also well fitted. In Fig.2 (a), the final classified groups are represented by red circle (light group) and blue circle (text group). Also Fig.3 shows that our method gives robust classification result even in the case when the linear assumption fails.

3. EXPERIMENTAL RESULTS

In this section, we present the results of experiments on two different types of images. Photos in Fig.1 are used as the test images. As shown in Fig.1 (a), the image is partially lightened and there are faint characters nearby 'patch' and 'α_f'. In the case of Fig.1 (b), it is so dark due to a shadow and the brightness is lower in the lower part. In order to show the ro-



(a) $\tau = 0.87$ (b) $\tau = 0.90$ (c) $\tau = 0.93$ (d) proposed algorithm



(e) $\tau = 0.87$ (f) $\tau = 0.90$ (g) $\tau = 0.93$ (h) proposed algorithm

Fig. 4. Comparison of retinex based algorithm and proposed algorithm

bustness of our algorithm, we compare with a retinex-based method which works on various images very well [7]. The results are shown in Fig.4 where the ratio threshold τ of a retinex-based algorithm varies from 0.87 to 0.93. It can be seen that a threshold of 0.93 matches well to Fig.1 (a) while this threshold to the Fig.1 (b) makes annoying noise. The threshold rather needs to be tuned to 0.87 as shown in Fig.4 (e). On the other hand, as shown in Fig.4 (d) and (h), our algorithm gives clear results for both cases without tuning parameters.

4. CONCLUSIONS

In this paper, we have introduced a new approach to the binarization of document images based on the energy minimization. The proposed algorithm is a soft decision method that do not need parameter control, whereas the previous algorithms need to control some parameters for thresholding. For energy minimization, we formulate the energy using MAP-MRF approach and perform the optimization via graph cut. Experimental results shows that our algorithm gives robust results to various types of images.

5. REFERENCES

[1] Z. Shi and V. Govindaraju, “Historical document image enhancement using background light intensity normalization,“ *ICDAR*, pp. 473-476, 2004.

[2] M. Ramirez, E. Tapia, M. block and R. Rojas, “Quantile linear algorithm for robust binarization of digitalized letters,“ *ICDAR*, pp. 1158-1162, 2007.

[3] Y. Li, C. Suen and M. Cheriet, “A threshold selection method mased on multiscale and graylevel co-occurence matrix analysis,“ *ICDAR*, pp. 575-579, 2005.

[4] Y. Xi, Y. Chen and Q. Liao, “A novel binarization system for degraded document images,“ *ICDAR*, pp. 287-291, 2007.

[5] W. Niblack, “An introduction to image processing,“ pp. 115-116, Prentice Hall, Englewood Cliffs, NJ, 1986.

[6] P. W. Palumbo, P. Swaminathan and S. N. Srihari, “Document image binarization : Evaluation of algorithms“, *Proc. SPIE*, pp. 278-286, 1986.

[7] M. Pilu and S. Pollard, “A light-weight text image processing method for handheld embedded cameras“, *bmvc*, Mar., 2002.

[8] Y. Y. Boykov and M. P. Jolly, “interactive graph cuts for optimal boundary & region segmentation of objects in N-D images,“ *Int. Conf. on Computer Vision*, vol. 1, pp. 105-112, July, 2001.