*ICIC Express Letters*
*Part B: Applications*
Volume **7**, Number **8**, August **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **1769–1774**

# IMPROVING KEYWORD SPOTTING ON DEGRADED HISTORICAL MONGOLIAN DOCUMENT IMAGES USING MARKOV RANDOM FIELD FOR RESTORATION

Hongxi Wei and Guanglai Gao

School of Computer Science
Inner Mongolia University
No. 235, Daxue West Road, Hohhot 010021, P. R. China
{ cswhx; csggl }@imu.edu.cn

Received December 2015; accepted March 2016

Abstract. *Due to aging, the scanned images of historical Mongolian document are degraded. In order to realize keyword spotting, the corresponding word images are segmented from the degraded document images. However, the problem of rupture and lack of stroke results in decreasing the performance of keyword spotting. In this paper, an approach based on Markov Random Field has been applied to improve the quality of the degraded word images. Each degraded word image is modeled by a Markov Random Field, in which the prior probability of the hidden-layer can be obtained by a codebook. The codebook is formulated by a training set of high quality binary word images. And the probability density of the observation-layer can be estimated on the global threshold of the input gray-level word images. In this way, the degraded gray-level word images can be converted into binary word images with better quality. The experimental results show that the Markov Random Field model can reduce degradation of word images so as to improve the performance of keyword spotting.*
**Keywords:** Historical Mongolian document, Document image retrieval, Degradation, Markov Random Field, Keyword spotting

1. **Introduction.** In the field of document image retrieval, keyword spotting technology is an alternative way when optical character recognition (OCR) is poor or hard. Keyword spotting was firstly introduced by Manmatha et al. [1] for indexing George Washington's manuscripts, which can locate keywords directly on document images by word matching.

In our previous work [2], an effective binarization approach has been proposed for the historical Mongolian document images. The detailed procedure is as follows. Firstly, the scanned color images are converted into gray-level images and smoothed by Wiener filter. Then, the gray-level images are processed by three well-known global thresholding methods, respectively. The three global methods are Otsu algorithm [3], Kittler algorithm [4] and fuzzy c-means clustering method (FCM) [5]. For a gray-level document image, the final binary image can be formulated by voting the three intermediate results. The corresponding binary word images can be segmented from each binary document image by connected component analysis. In this way, a collection of word images can be obtained from the corresponding collection of document images.

For accomplishing keyword spotting on historical Mongolian document images, each binary word image should be represented by several kinds of profile-based features (such as project profile, left profile, right profile, and background-to-foreground transitions) [6]. Furthermore, a fixed-length feature vector is formulated by obtaining the appropriate number of the complex coefficients of discrete Fourier transform on each profile feature [6,7]. In this way, online image-to-image matching can be supported by calculating similarities (such as Euclidean distance) between a query keyword and each word image in the collection.

However, the binary word images have much more degradations, such as holes and ruptures of stroke. Therefore, our motivation is to promote the quality of word images so that improve keyword spotting on historical Mongolian document images.

In literature [8,9], the Markov Random Field (MRF) model has been applied in image restoration and especially suited for historical or handwritten document images. In this paper, a patch-based MRF model has been used to restore word images of historical Mongolian document. Each word image is divided into a set of patches with equal sizes, in which non-neighboring patches are independent. Only neighboring patches are dependent on each other. This kind of dependent relationship is modeled by the MRF model. In the MRF model, the prior probability of the hidden-layer can be obtained by a codebook. The codebook is formulated by a training set of high quality binary word images. And the probability density of the observation-layer can be estimated on the global threshold of the input gray-level word images. The details will be described in the following sections.

The rest of the paper is organized as follows. Section 2 explains the MRF model for image restoration in detail. The experimental results are shown in Section 3. The conclusive remarks are presented in Section 4.

## 2. MRF Based Image Restoration.

A patch-based MRF model is used to restore word images of historical Mongolian document, which is similar to [9]. In the patch-based MRF model, each gray-level word image is divided into a number of non-overlapping square patches (such as $g_1$, $g_2$, ..., $g_N$) with the same sizes. And the corresponding binary word image is also divided into patches (such as $b_1$, $b_2$, ..., $b_N$). Each gray patch $g_i$ $(1 \le i \le N)$ only depends on its corresponding binary patch $b_i$. Each binary patch $b_i$ $(1 \le i \le N)$ conditionally depends on its four neighboring patches in both the horizontal and vertical directions. Therefore, the conditional probability formulas are as follows:

$$P(g_i|b_1, b_2, \ldots, b_N, g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_N) = P(g_i|b_i),\ 1 \le i \le N \tag{1}$$

$$\begin{aligned} &P(b_i|b_1, b_2, \ldots, b_{i-1}, b_{i+1}, \ldots, b_N, g_1, g_2, \ldots, g_N) \\ &= P(b_i|b_{i,n1}, b_{i,n2}, b_{i,n3}, b_{i,n4}),\ 1 \le i \le N \end{aligned} \tag{2}$$

where $b_{i,n1}$, $b_{i,n2}$, $b_{i,n3}$, and $b_{i,n4}$ are the four neighboring patches of $b_i$.

In the patch-based MRF model, the binary word image can be estimated by calculating maximum a posteriori (MAP) by the following equation.

$$P(b\,|g) = P(b, g)/P(g) \tag{3}$$

where $g$ is an input (or observed) gray-level word image, $b$ is the estimated (or inferred) binary word image, and $P(.)$ is a function of probability density. In (3), the denominator is a constant for $b$. Thus, only the joint probability $P(b, g)$ needs to be computed. And in the MAP approach, each patch $b_i$ $(1 \le i \le N)$ of the inferred binary word image can be estimated by (4) using the joint probability $P(b, g)$.

$$\hat{b}_i = \underset{b_i}{\operatorname{argmax}} \max_{b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_N} P(b, g) \tag{4}$$

However, it is impossible to calculate (4) directly for large images because the computation grows exponentially as the number of patches increases [9]. The belief propagation (BP) algorithm can be approximated the MAP estimation in linear time according to the number of patches. The BP algorithm will be introduced detailedly in the following subsection.

### 2.1. Belief propagation.

The joint probability $P(b, g)$ can be computed fast by the BP algorithm. And the joint probability is expressed by the following formulation.

$$P(b_1, \ldots, b_N, g_1, \ldots, g_N) = \prod_{(i,j)} \varphi(b_i, b_j) \prod_k \phi(b_k, g_k) \tag{5}$$

where $\varphi(b_i, b_j)$ and $\phi(b_k, g_k)$ are the pairwise compatibility functions, which are learned from the training data; $(i, j)$ indicates neighboring patches between $b_i$ and $b_j$, and $N$ is the number of patches. In this way, Equation (4) can be rewritten as follows according to (5):

$$\hat{b}_i = \underset{b_i}{\operatorname{argmax}} \max_{b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_N} \prod_{(i,j)} \varphi(b_i, b_j) \prod_k \phi(b_k, g_k) \tag{6}$$

Utilizing the BP algorithm, Equation (6) can be computed by iterating the following steps. The MAP estimation at patch $b_i$ is

$$\hat{b}_i = \underset{b_i}{\operatorname{argmax}}\, \phi(b_i, g_i) \prod_k M_i^k \tag{7}$$

where $k$ runs over all neighboring patches of patch $b_i$, and $M_i^k$ is the message from patch $b_i$ to patch $b_k$. And $M_i^k$ is calculated by (8).

$$M_i^k = \max_{[b_k]} \varphi(b_i, b_k)\phi(b_k, g_k) \prod_{l \neq i} \tilde{M}_k^l \tag{8}$$

where $\tilde{M}_k^l$ is $M_k^l$ from the previous iteration. The initial $\tilde{M}_k^l$'s are set to column vectors of $l$'s with the dimension of the patch $b_i$.

In (8), the compatibility functions $\varphi(b_i, b_k)$ and $\phi(b_k, g_k)$ are defined as follows.

$$\varphi(b_i, b_k) = \frac{P(b_i, b_k)}{P(b_i)P(b_k)} = \frac{P(b_i|b_k)}{P(b_k)} = \frac{P(b_k|b_i)}{P(b_i)} \tag{9}$$

$$\phi(b_k, g_k) = P(b_k, g_k) = P(b_k)P(g_k|b_k) \tag{10}$$

Thus, we can obtain (11) and (12).

$$\hat{b}_i = \underset{b_i}{\operatorname{argmax}}\, P(b_i)P(g_i|b_i) \prod_k M_i^k \tag{11}$$

$$M_i^k = \max_{[b_k]} P(b_k|b_i)P(g_k|b_k) \prod_{l \neq i} \tilde{M}_k^l \tag{12}$$

Here, only the prior probabilities $P(b_i)$, $P(b_k|b_i)$ and the observation probability $P(g_i|b_i)$ (or $P(g_k|b_k)$) need to be estimated in advance for using (11) and (12). In the next two subsections, the procedures of estimating the prior probabilities and the observation probability will be introduced, respectively.

2.2. **Estimating prior probability.** In order to estimate the prior probabilities $P(b_i)$ and $P(b_k|b_i)$, a set of high-quality binary word images has been collected in advance. There are 5500 binary word images in total and they are obtained using the approach proposed in [2]. Each high-quality binary word image is divided into a number of non-overlapping patches. The size of one patch is $15 * 15$ determined by the thickness of the stroke. Because the binary patches may be similar to each other, the k-means algorithm is used to form several representatives of the binary patches.

In this study, 233 code words (i.e., representatives) were determined by the k-means algorithm, which was formulated the codebook. Figure 1 lists the 233 code words of the codebook. So, the prior probability $P(b_i)$ is estimated by the following way.

$$P(b_i = C_l) = n/N,\ l = 1, 2, \ldots, M \tag{13}$$

where $M$ is the number of the code words (i.e., $M = 233$), $C_l$ is the $l^{\text{th}}$ code word, $n$ is the number of patches most similar to $C_l$, and $N$ is the total amount of the binary patches in the training set. The sum of $P(b_i)$ equals to one.

Based on the Bayesian theory, the prior probability $P(b_k|b_i)$ can be computed by

$$P(b_k|b_i) = P(b_i, b_k)/P(b_i) \tag{14}$$

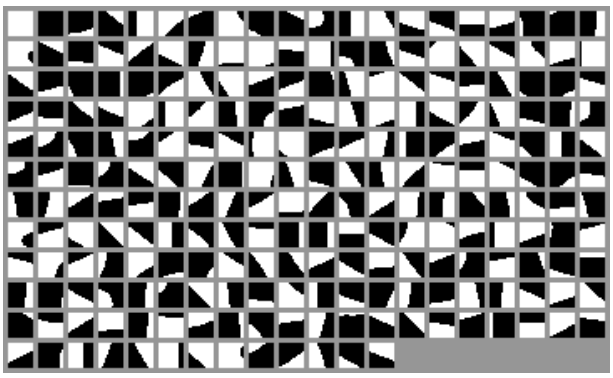

FIGURE 1. The 233 code words of the codebook

In Equation (14), the probability $P(b_i)$ is estimated using (13) and the joint probability $P(b_i, b_k)$ needs to be estimated in the horizontal and vertical directions, severally. The joint probability of the horizontal direction is estimated by

$$P(b_i = C_l, b_k = C_m) = n_h/N_h \tag{15}$$

where $l = 1, 2, \ldots, M$, $m = 1, 2, \ldots, M$; $C_l$ and $C_m$ are the $l^{\text{th}}$ and $m^{\text{th}}$ code words, separately; $n_h$ is the number of the neighboring patches in the horizontal direction, in which the left patch is most similar to $C_l$ and the right one is most similar to $C_m$; $N_h$ is the total amount of the neighboring binary patches in the horizontal direction over the training set. In the same manner, the joint probability of the vertical direction can be also estimated.

2.3. **Estimating observation probability.** On the patch level, the pixels of the gray patch and the corresponding binary patch still obey the conditional dependence assumption. Therefore, the observation probability is estimated by

$$P(g_i|b_i) = P\left(g_i^{1,1}, \ldots, g_i^{S,S} \middle| b_i^{1,1}, \ldots, b_i^{S,S}\right) = \prod_{r=1}^{S} \prod_{c=1}^{S} P\left(g_i^{r,c}|b_i^{r,c}\right) \tag{16}$$

where $g_i^{r,c} (1 \le r, c \le S)$ is the pixel in row $r$ and column $c$ of $g_i$, and $S$ is the rows (or columns) of a squared patch.

Suppose the distribution of the intensity of foreground is $pdf_f(g_i^{r,c}) = P(g_i^{r,c}|b_i^{r,c} = 1)$ and the distribution of the intensity of background is $pdf_b(g_i^{r,c}) = P(g_i^{r,c}|b_i^{r,c} = 0)$.

Hence, the observation probability is calculated by

$$P(g_i|b_i) = \prod_{1 \le r,c \le S}^{b_i^{r,c}=1} pdf_f\left(g_i^{r,c}\right) \prod_{1 \le r,c \le S}^{b_i^{r,c}=0} pdf_b\left(g_i^{r,c}\right) \tag{17}$$

Assume that $pdf_f$ and $pdf_b$ are two normal distributions. To determine the means and variances of the foreground and background, for each gray-level word image, a global threshold is obtained using the Otsu algorithm [3]. Thus, the means and variances can be calculated by the foreground and background pixels, respectively. These means and variances would be used in Equation (17) for obtaining the observation probability of each gray-level word image.

3. **Experimental Results.** To evaluate the performance of MRF, a data set is collected which consists of 200 scanned Mongolian Kanjur images. And each Kanjur image has been transcribed manually by Unicode to form the ground truth data.

By analyzing the ground truth data, twenty meaningful words are selected and taken as query keywords in the experiment. And for each query keyword, the number of occurrences is more than twenty times. Both the data set and the twenty query keywords are the same as in [7]. The evaluation metric is *R-Precision*, and its formula is as follows:

$$R\text{-}Precision = r/Rel \tag{18}$$

where $r$ is the number of relevant results in the top of *Rel* returned results. For each query keyword, a *R-Precision* can be calculated. Thus, the *R-Precision* of all query keywords are averaged by the amount of query keywords, which is taken as the performance measure.

In our experiment, 49444 gray-level word images are segmented from the corresponding dataset, which are considered as a collection of retrieval objects. For comparison, these gray-level word images are converted into binary images by voting approach [2] and MRF, separately. In [7], the average *R-Precision* of the twenty query keywords is **60.27**%, which is taken as a baseline. The average *R-Precision* of MRF is **63.39**%. The performance of MRF is better than the baseline, which has been increased by **3.12**%.

Figure 2 gives the comparison results between MRF and the previous voting approach [2]. Figure 3 shows the individual *R-Precision* of each query keyword. It comes to a conclusion that MRF is superior to [2] for almost all query keywords. The *R-Precision* of some query keywords with or without MRF are equivalent, which indicates that the original approach can obtain the same quality as MRF only for a part of word images.

FIGURE 2. The comparison results between MRF and the original voting approach

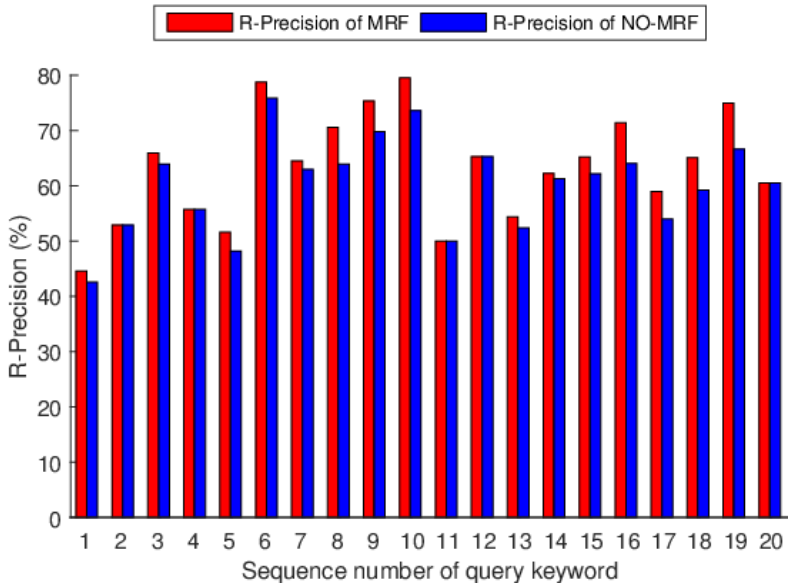


FIGURE 3. The *R-Precision* of each query keyword

4. **Conclusion.** In this paper, a patch-based MRF model has been adopted to reduce the degradations of word images. In the MRF model, the prior probability of the binary patches is estimated by training a codebook. The code words are determined by the k-means algorithm. Utilizing the codebook, the prior probabilities between two neighboring binary patches are estimated in the horizontal and vertical directions, respectively. Moreover, suppose the distributions of the intensity of foreground and background are two normal distributions. For a gray-level word image, each pixel can be classified into

foreground and background by its global threshold. And then, the means and variances of the two normal distributions can be calculated. Combining other approaches with the MRF model to improve the quality of word images is the future work of our research.

**Acknowledgement.** This work is supported by the National Natural Science Foundation of China under Grant 61463038 and the Research Project of Higher Education School of Inner Mongolia Autonomous Region of China under Grant NJZY14007.

## REFERENCES

[1] R. Manmatha, C. Han, E. M. Riseman and W. B. Croft, Indexing handwriting using word matching, *Proc. of the 1st ACM International Conference on Digital Libraries (ICDL)*, Bethesda, United States, pp.151-159, 1996.

[2] H. Wei, G. Gao, Y. Bao and Y. Wang, An effective binarization method for ancient Mongolian document images, *Proc. of the 3rd International Conference on Advanced Computer Theory and Engineering*, Chengdu, China, pp.43-46, 2010.

[3] N. Otsu, A threshold selection method from gray level histograms, *IEEE Trans. Systems Man Cybernetics*, vol.9, no.1, pp.62-66, 1979.

[4] J. Kittler and J. Illingworth, Minimum error thresholding, *Pattern Recognition*, vol.19, no.1, pp.41-47, 1986.

[5] R. Duda, P. Hart and G. David, *Pattern Classification*, 2nd Edition, Wiley, New York, 2001.

[6] H. Wei, G. Gao and X. Zhang, Indexing for Mongolian Kanjur images in word spotting, *Journal of Computational Information Systems*, vol.9, no.4, pp.1501-1508, 2013.

[7] H. Wei and G. Gao, A keyword retrieval system for historical Mongolian document images, *International Journal on Document Analysis and Recognition*, vol.17, no.1, pp.33-45, 2014.

[8] J. P. Kuk, N. I. Cho and K. M. Lee, MAP-MRF approach for binarization of degraded document image, *Proc. of the 15th International Conference on Image Processing (ICIP)*, San Diego, United States, pp.2612-2615, 2008.

[9] H. Cao and V. Govindaraju, Preprocessing of low-quality handwritten documents using markov random fields, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.7, pp.1184-1194, 2009.