International Journal of Advanced Smart Convergence Vol.4 No.2 170-176 (2015)
http://dx.doi.org/10.7236/IJASC.2015.4.2.170

IJASC 15-2-24

Mongolian Traditional Stamp Recognition using Scalable kNN

Gantuya. P*, Mungunshagai. B† and Suvdaa. B*

*School of Engineering and Applied Sciences, National University of Mongolia
†Fine Arts Design and Technology School, Ikh Zasag International University
E-mail: {gantuya, suvdaa}@seas.num.edu.mn, listat0807@gmail.com

Abstract

The stamp is one of the crucial information of traditional historical and cultural for nations. In this paper, we purpose to detect official stamps from scanned document and recognize the Mongolian traditional, historical stamps. Therefore we performed following steps: first, we detect official stamps from scanned document based on red-color segmentation and document standard. Then we collected 234 traditional stamp images with 6 classes and 100 official stamp images from scanned document images. Also we implemented the processing algorithms for noise removing, resize and reshape etc. Finally, we proposed a new scale invariant classification algorithm based on KNN (k-nearest neighbor). In the experimental result, our proposed a method had shown proper recognition rate.

Keywords: *Scanned Document Correction, Stamp Detection, Color Based Segmentation Stamp Recognition, Image Processing, KNN*

1. INTRODUCTION

Nowadays, when computer technology is present in various areas of life, the problem of computer crime is becoming more and more important. It covers both strictly electronic and traditional types of law-breaking. On the other hand, there are still many areas of life, where computers and digital media are employed only as tools and play just a supporting role. The most evident example of such domain is an area associated with official documents, identity cards, formal letters, certificates, etc. All these documents are being issued by formal authorities and are often in a form of a paper letter consisting of several typical elements: heading, body text, signatures and stamps which, from this historical point of view confirm its official character. In business environments, they are often used to provide supplemental information. In other words, its main purpose is to authenticate a document which in many cases is a subject to forgery or tampering with help of modern computer means. In general, the process of forgery consists of the following steps: obtaining the original document, high resolution scanning, digital image manipulation and final printing. It is rather easy to recognize fake stamps, even if they are printed using ink-jet printers.

On the other hand, reliable recognizing stamps in the documents is not trivial and has not been solved so far The most advanced method found in the scientific literature is described in, where the authors present a stamp detection approach, which treats stamps as regions with analytically shaped contours, however, these

Manuscript Received: Sept. 18, 2015 / Revised: Oct. 15, 2015 / Accepted: Nov. 10, 2015
Corresponding Author: gantuya@seas.num.edu.mn
Tel: (+01-976-11) 32-40-07
School of Engineering and Applied Sciences, National University of Mongolia

regions are limited to oval shapes only. [6–8].
The general motivation of the research presented in this paper is Mongolian documentation standard that is able to analyze an image as well as detect and localize different types of stamps in it.[1] The application area of this kind of a system is broad, law-enforcement forces, law offices, official archives and any other stamp utilizing institutions.

2. RELATED WORKS

The cultural behavior, farming, religion, script and fine arts which the ancient Mongolians used to run are expressed as stamp shape. The stamp is significant for defining ancestry, symbol of tribes and is significant for defining the derivation of historical findings. These are considered to be divided into four things such as pitch-brand, stamps on rocks and tribe stamp. Of the above, the tribe stamp is the most important symbol. [4] We have intended to create database with the image of the tribe stamp and recognize it [1] [2]. By researching the tribe stamp, we can create the training database and classify it using a new algorithm.
There are many historical researches that classify the stamps [3] divided into basic and derivative ones according to their founded territories. In this paper we classify them based on their shapes using kNN based algorithm.[5] We have investigated the total of 931 stamps using their basic images.[3] In this paper, we used 234 of them.

In general there is no specific research that Mongolian stamp recognition using any algorithm from stamp image. Therefore we reviewed the following research works that some sign image recognition works. The KNN algorithm is a popular method for pattern recognition, it has proposed several studies [5], [9], [10]. Also the stamp and sign recognition studies are introduced following researches. Traffic sign detection and recognition [14], Postage stamp recognition using image processing [11], Photo time-stamp detection and recognition [12] and Text detection in images based on unsupervised classification of edge-based features [13].

3. PAPER TITLE AND AUTHOR INFORMATION

We first do preprocessing for scanned document. And then we detect red colored part in the bottom part of the image and crop original image with detected location mask. Then we do resize and reshape the cropped different sized rectangle images into same dimension vectors. For feature extraction step, we use Principal Component Analysis (PCA). Finally, we classify the stamp images by KNN algorithms. Figure 1 has shown the main structure of our work.

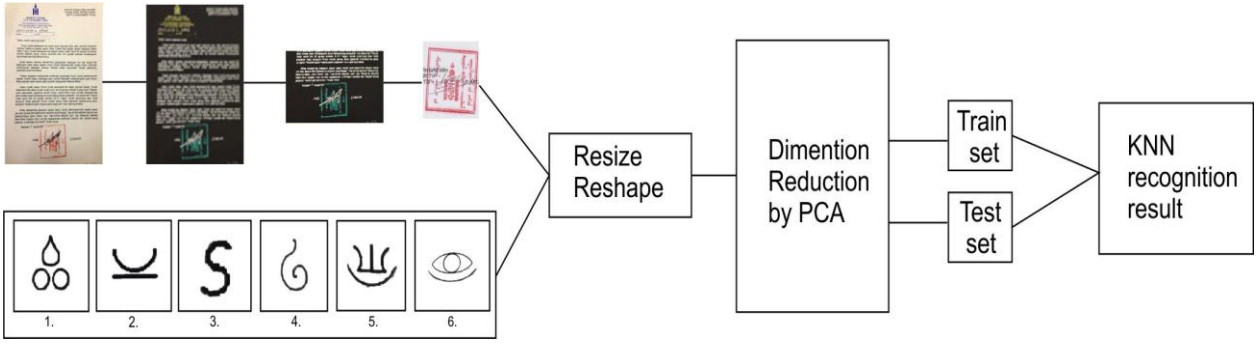

Figure 1. Structure of this work.

3.1. Stamp detection based on Mongolian stamp standard for official documents

According to part 5.1-6 of Mongolian official document structures and its requirements standards, the stamp must be located in the bottom part of page and overlaid with signature. [1-4] We propose a method based on the standard. Official stamps are classified several classes by their shapes and organizations. In this part, we detect stamps based on red-color segmentation and location of the bottom place.

Also we can classify stamps by organization. For example non-governmental organization's stamp and governmental organization's stamp etc. Figure 2 has shown an example of non-governmental (right) and governmental (left) scanned document.

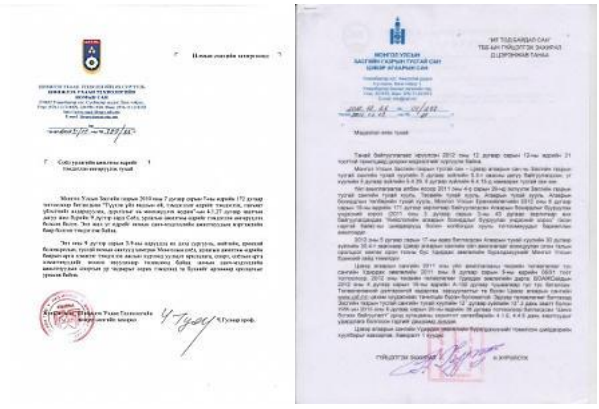

Figure 2. An examples of non-governmental (right) and governmental (left) stamps

The governmental stamps are rectangular shaped. The non-governmental organization's stamps are different shaped.

Threshold

The algorithm of processing is divided into two parts. Its main purpose is to obtain color information related to the specific documents and stamps being used. An input image has to be stored in a file with possibly lossless compression, high spatial resolution and full color range (24bit RGB). For the most popular red colored stamps, the RGB image is thresholded by its red color value.

Figure 3. Classes of stamps in view point of shapes

As it can be seen from the presented examples, round-shaped stamps are much more similar to each other than rectangular ones. (Figure 3.)

3.2. Dimension Reduction

The Principal Component Analysis (PCA) examines relationships of variables. It can be used to reduce the number of variables in regression and classification. Each principal component in PCA is the linear combination of the variables and gives a maximized variance. Let *X* be a matrix for *n* observations by *p* variables, and the covariance matrix is *S*. PCA has following 4 steps:

1. Find mean_subtracted_vectors by computing mean vector among n observations and subtract each vector from mean vector
2. Compute covariance matrix S_{pxp} from above step's mean_subtracted_vectors.
3. Find eigenvalues and eigenvectors from the covariance matrix. Select k eigenvectors (principal components) with the highest k eigenvalues.
4. Project original data by selected eigenvectors' matrix. $Y_{nxk}=X_{nxp}*G_{pxk}$

Where G is eigenvectors' matrix p by k, Y is projected data matrix n by k and p>>k.
PCA reduces dimension from p to k.

3.3. Stamp classification

The K-Nearest Neighbor (KNN) algorithm is a method for classifying objects based on the closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The KNN algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors by Euclidean distance metric, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of its nearest neighbor.

The proposed method is scale invariant method based on KNN. The original KNN algorithm uses Euclidean distance for comparing the test image and training image from training database. The Euclidean distance is rotation and translation invariant. But it is not scale invariant. Therefore some case, the size of stamp is changed, it fails. Then we propose scale invariant KNN algorithm using distance ratio method as following equation (1).

$$R_{i+1}= d(x_{i+1}, x_{i+2})/d(x_i, x_{i+1}) \qquad (1)$$

Where $d(x_{i+1}, x_{i+2})$ is Euclidean distance between $(x+1)^{th}$ vector and $(x+2)^{th}$ vectors and R_{i+1} is the ratio of those features. This distance is computed for each followed two features.

4. EXPERIMENTAL RESULTS

4.1. Database

We collected 100 official stamp images with 2 classes and 234 traditional stamp images with 6 classes.

Stamp of class:

1. Chandman stamps
2. Saran stamps
3. Zuuzai stamps
4. Erdnes stamps
5. Bagaj zevseg stamps
6. Hun amitan stamps

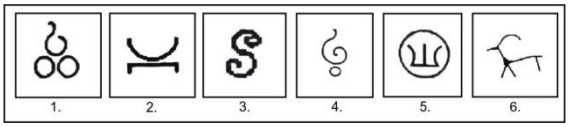

Figure 4. Image traditional stamp

4.2. Stamp detection results

We have done experiments in 100 images using Matlab.

Input image format was RGB. Figure 4-5 has shown an example of input image.

Figure 5. An input image

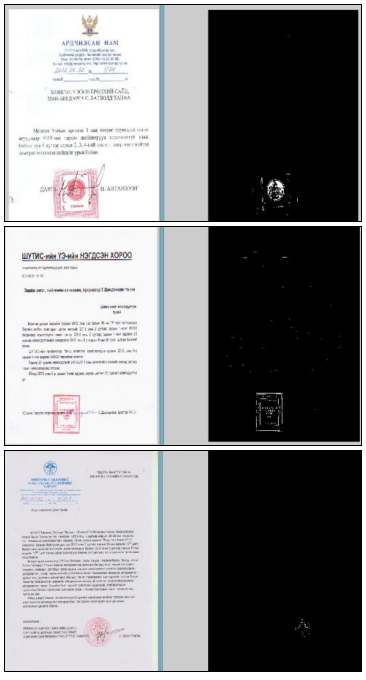

Figure 6. Stamp detection result

Figure 6 has shown the result of the thresholding by red color.

Figure 7. The cropped stamps

Figure 7 has shown the result of the cropped images. The detection result was 97% among the collected 100 images.
After cropping the stamps, we collect them and prepare training dataset for two classes by shapes (round and rectangular).

4.3. Stamp classification results

Official stamp recognition
In this part, we classify the stamps that one of the stamps was recognized as the round type, while the other was classified as the rectangular one.
The cropped images are scaled to size 80 by 80 pixels, so the feature vector size becomes 6400. Next step, the PCA is reduced dimension of feature vector to 50, 80, 100, 150, … dimensions. After doing several experiments, we select 100 (the optimal size) dimension that recognition rate is highest.
Knn classifies the stamps by their shapes into two classes. Training and test data ratio was 70:30 and the recognition rate was 86.46±5.74.

Traditional Stamp recognition
Using the algorithm KNN about 234 stamps images were divided into 6 groups. We test 120 images from training set and 114 images from test set (not in training set). The recognition rate has shown in Table 1.

Table 1. Comparison of recognition rate of two methods

		# of images	*Recognition rate (%) of original KNN*	*Recognition rate (%) of proposed method*
K=3	Train error	120	94	100
	Test error	114	91	96

The error of the original KNN depends on the scale variations. Therefore the new method could improve the recognition rate.

5. CONCLUSION

In this paper we propose a method to classify 2 kinds of stamp images that one is official stamps and the other traditional stamps. We detected the official stamps based on red color and its position. And then we classify stamps by shapes using PCA and KNN algorithms. Our experimental results were reasonable. In future, it is possible to verification of the governmental documents based on the stamp.

REFERENCES

[1] Mongolian National Standard, *Documents*, Mongolian Agency for Standardization and Metrology, 2002.

[2] Ts.Battulga, *Mongolian Runi documents*, 2005.

[3] Kh.Perlee, *Research collection*, 2012.

[4] S.Dulam, *Symbolization*, 2010.

[5] Hossein Ebrahimpour and Abbas Kouzani, "Face Recognition using bagging KNN," Int'l Conf. on CVPR, IEEE, pp. 209-216, 1996.

[6] Pawel Forczmanski, *Stamp detection in scanned documents*, Chair of Multimedia Systems, West Pomeranian University of Technology, pp.61-68., 2010.

[7] Barbora Micenkovfi, Joost van Beusekom, and Faisal Shafait, *Stamp Verification for Automated Document Authentication*, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany, 2009.

[8] Guangyu Zhu*, Stefan Jaeger, David Doermann, *A robust stamp detection framework on degraded documents*, Institute for Advanced Computer Studies, University of Maryland, College Park, MD, USA 20742, 2005.

[9] Patrik K, Martina Z, Robert H, Roman J, Miroslav B and Jan H, *A Novel Approach to Face Recognition using Image Segmentation Based on SPCA-KNN Method*, University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia, 2013.

[10] Mohammad K. H., Abu A. S. R., Rajibul A., and Dr.William P., *Automatic Face Recognition System using P-tree and K-Nearest Neighbor Classifier*, 2004.

[11] Guancong LiJune, "Postage stamp recognition using image processing", Faculty of Engineering and Sustainable Development, 2011.

[12] Xiang-rong Chen and HongJiang Zhang, "Photo time-stamp detection and recognition," 7th International Conference on Document Analysis and Recognition (ICDAR 2003), 2-Volume Set, Edinburgh, Scotland, UK, pp.319-322, 2003.

[13] Chunmei Liu, Chunheng Wang and Ruwei Dai, "Text detection in images based on unsupervised classification of edge-based features," In proceeding of: Document Analysis and Recognition, Seoul, Korea, pp. 610–614, August 2005.

[14] Hasan Fleyeh, *Traffic and Road Sign Recognition*, doctoral thesis, Napier University, 2008.

[15] Usman Zakir, *Automatic Road Sign Detection and Recognition*, doctoral thesis, Department of Computer Science, Loughborough University, 2011.