International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:4, 2007

Support Vector Machine for Persian Font Recognition

A. Borji, and M. Hamidi

***Abstract*—**In this paper we examine the use of global texture analysis based approaches for the purpose of Persian font recognition in machine-printed document images. Most existing methods for font recognition make use of local typographical features and connected component analysis. However derivation of such features is not an easy task. Gabor filters are appropriate tools for texture analysis and are motivated by human visual system. Here we consider document images as textures and use Gabor filter responses for identifying the fonts. The method is content independent and involves no local feature analysis. Two different classifiers Weighted Euclidean Distance and SVM are used for the purpose of classification. Experiments on seven different type faces and four font styles show average accuracy of 85% with WED and 82% with SVM classifier over typefaces

***Keywords*—**Persian font recognition, support vector machine, gabor filter.

I. Introduction

FONT recognition is a preliminary step in optical character recognition systems. Most previously proposed methods in literature to recognize font styles make use of local typography features [1..5]. On the other side are methods that use global information instead of local features [6..10]. One benefit of global methods is their applicability over many different domains and eliminating the burden of developing specific local features for each problem [11]. Although global texture analysis methods have shown great performance for English language, their applicability over other languages is yet under question [12]. Here we are going to examine the applicability of Gabor filters for the purpose of Persian font recognition. Figure one shows an overall sketch of our font recognition system.

A set of training samples each one a different font image is fed to normalization component. Then Gabor filters are applied to normalized images and features are extracted. A classifier is trained on its input feature vectors. After training, classifier could be used to identify an image font in test phase.

The rest of the paper is organized as follows: in section 2, normalization of text images is discussed. Feature extraction using multichannel Gabor filtering is illustrated in section 3. Section 4 outlines the classifiers. Experiments and Results are discussed in section 5. Finally Conclusions and future works are drawn in section 6.

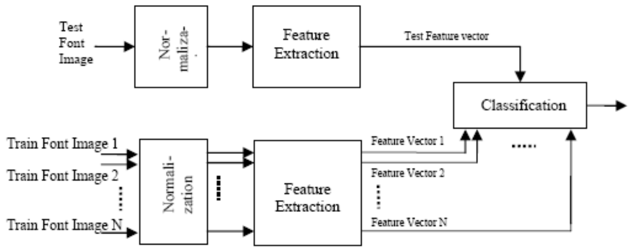

Fig. 1 Overall structure of font recognition system

II. Normalization of Document Images

The original image could not be used directly. It must be normalized according to spaces between lines and individual words. Normalization steps are as follows:

1) Use Otsu algorithm on original image to obtain a binary image [13].
2) Compute the HPP1 of the image. Remove zeros from this profile and smooth it with averaging. The peaks and valleys correspond to line centers and space between them respectively.
3) For each text line compute the VPP2. The valleys between peaks correspond to spaces between words. Normalize the spacing by scaling them to a predefined width. In the case of incomplete lines, the spaces should be filled up by repeating the line until it reaches a predefined length.
4) If document still contains spacing, fill it up by repeating the first lines to obtain an image with a predefined size (Here 300 * 300). Resultant image is divided to 25, 100*100 non-overlapping blocks. Following figure shows an example after performing above steps on a sample image.

Ali Borji is now a phd student at the School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran (corresponding author e-mail: borji@ipm.ir).
Mandana Hamidi is an MSc. Student at Islamic Azad Univeristy of Qazvin. She is also with the faculty of Islamic Azad University of Zarghan, Zarghan, Iran (e-mail: Mandana.Hamidi@gmail.com).

[1] Horizontal projection profile
[2] Vertical projection profile

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:4, 2007

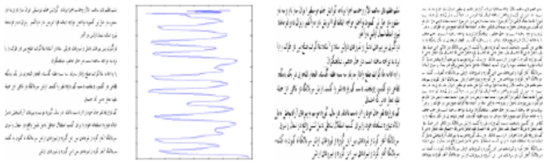

Fig. 2 Example of Normalization.(Original image, HPP, normalized height image, final normalized image)

III. Feature Extraction

Once uniform blocks of text created, we can proceed with feature extraction based on texture analysis. Two methods we use are multichannel Gabor filtering and Gray scale Co-occurrence Matrices. A brief explanation of these methods comes below.

A. Gabor Filters

Simple cell responses in area V1 of visual cortex have main role in edge detection and texture processing. They can be modeled as Gabor filters of the stimulus, parameterized by location, orientation, scale and phase [14] as in Fig. 3. A simple model of these cells is described in [15]. A 2D Gabor filter $g(x, y)$ can be formulated as:

$$
\begin{aligned}
&g_{\gamma,\lambda,\theta,\varphi}(x,y)=\exp(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2})\cos(2\pi\frac{x'}{\lambda}+\varphi)\\
&x'=x\cos\theta-y\sin\theta\\
&y'=x\sin\theta-y\cos\theta
\end{aligned} \tag{1}
$$

In above equations, x and y represent image coordinates, σ is the standard deviation of Gaussian function which is usually set to 0.56 λ, λ is the wave length of cosine equation, γ characterizes the shape of Gaussian, circular shape for γ=1 and elliptic for γ<1, θ represents the channel orientation and takes values in interval [0 π], φ is the phase of cosine function and is symmetric for values 0 and π.

Fig. 3 A 2D Gabor function Left: Frequency domain Right: Spatial domain

For a given input image I(x,y) the response of Gabor filter is derived from the following convolution:

$$
r_{\gamma,\lambda,\theta,\varphi}(x,y)=\iint I(\varepsilon,\eta)\,g(x-\varepsilon,y-\eta)\,d\varepsilon\,d\eta \tag{2}
$$

Then the outputs of two symmetric and anti-symmetric filters (corresponding with φ=0 and φ= - π/2) are combined to provide a single channel output.

$$
E_{\gamma,\lambda,\theta}=({r_{\gamma,\lambda,\theta,0}}^2+{r_{\gamma,\lambda,\theta,-\pi/2}}^2)^{\frac{1}{2}} \tag{3}
$$

In above equation E is known as the energy of Gabor filter.

There are some important considerations in selecting the channel parameters f, θ, and σ. In [15], it is showed that there is no need to uniformly cover the entire frequency plane so far as texture recognition is concerned. Since the Gabor filters we use are of central symmetry in the frequency domain, only half of the frequency plane is needed. Eight values of orientation θ are used: θ = k(π/8), k=1,2…8. For each orientation, we use (λ =2.7, 4.1, 5.4) as spatial frequencies. This gives a total of 24 Gabor channels (four orientations combined with three frequencies). The above choice is sufficient to discriminate different fonts. The spatial constants σ of these channels, which determine the channel bandwidths, are chosen to be inversely proportional to the central frequencies of the channels [15]. Frequency responses of the Gabor filters used to identify different fonts are shown in Fig. 4.

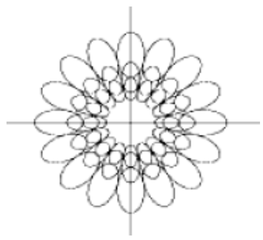

Fig. 4 Frequency responses of Gabor filters used in font identification. There are a total of 24 Gabor channels

IV. Font Recognition

We used two different classifiers, Weighted Euclidean Distance (WED) and Support Vector Machines (SVM) for the purpose of font recognition. In principle, any type of classifiers could be used. Here a short description of these classifiers is considered.

A. Weighted Euclidean Distance Classifier (WED)

In order to classify a handwritten text, features are extracted after normalization and filtering, then the feature vector is compared against all feature vectors of all writers in the training set using weighted Euclidean distance. The writer of handwritten text is k which minimizes the following distance function:

$$
d(k)=\sum_{i=1}^{N}\frac{(f_i-f_i^k)^2}{(\delta_i^k)^2} \tag{4}
$$

Where f_i is the ith feature of feature vector, f_i^k and δ_i^k are the mean and standard deviation of the ith feature of writer k over blocks respectively.

B. Support Vector Machines Classifier (SVM)

SVMs belong to the class of maximum margin classifiers. They perform pattern recognition between two classes by finding a decision surface that has maximum distance to the closest points in the training set which are termed support vectors. We start with a training set of points $x_i \varepsilon IR^n$, i = 1, 2, …, N where each point xii belongs to one of two classes identified by the label y_i Є {-1, 1}. Assuming linearly separable data3, the goal of maximum margin classification is to separate the two classes by a hyperplane such that the distance to the support vectors is maximized. This hyperplane is called the optimal separating hyperplane (OSH). The OSH has the form:

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i x_i.x+b, \tag{5}$$

The coefficients α_i and b the b in Eq. (1) are solutions of a quadratic programming problem [18]. Classification of a new data point x is performed by computing the sign of the right side of Eq. (1). In the following we will use:

$$d(x)=\frac{\sum_{i=1}^{l}\alpha_i y_i x_i.x+b}{\|\sum_{i=1}^{l}\alpha_i y_i x_i.x\|}, \tag{6}$$

to perform multi-class classification. The sign of d is the classification result for x and |d| is the distance from x to the hyperplane. Intuitively, the farther away a point is from the decision surface, i.e. the larger |d| the more reliable the classification result.

The entire construction can be extended to the case of nonlinear separating surfaces. Each point x in the input space is mapped to a point z = Φ(x) of a higher dimensional space, called the feature space, where the data are separated by a hyperplane. The key property in this construction is that the mapping Φ(.) is subject to the condition that the dot product of two points in the feature space Φ(x) . Φ(x) can be rewritten as a kernel function K(x,y). The decision surface has the equation:

$$f(x)=\sum_{i=1}^{l}\alpha_i\ y_i\ K(x_i,x)\ +b \tag{7}$$

again, the coefficients αi and b and are the solutions of a quadratic programming problem. Note that f(x) does not depend on the dimensionality of the feature space. In our experiments we use a family of kernel functions known as polynomial kernel:

$$K(x,y)=(1+x\cdot y)^d, \tag{8}$$

where d is the degree of the polynomial. In this case the components of the mapping Φ(x) are all the possible monomials of input components up to the degree d.

IV. Experimental Results

Several experiments are carried out to test the validity of Gabor filters for Persian font recognition. Seven frequently used Persian typefaces (Badr, Divani Mazar, Jadid, Titr, Siavash Mazar, Nasim, Homa) combined with four styles (Regular, Bold, Italic and Bold-Italic) are trained and tested. In other words a total of 28 fonts are used in our experiments.

For convenience we use computer generated images. The input image is normalized to form a 300 * 300 text image as mentioned in section 2. Then it is divided to 9 non-overlapping blocks, each of size of 100*100. For each font, we randomly chose 3 such blocks for training and remaining 6 blocks for testing. For each block, all 24 filters are applied and then another image is derived by taking the maximum of these filter responses per pixel. The mean values (M) and the standard deviations (S) of the channel output images (over each block) are chosen to represent texture features. Thus, a total of 50 features are extracted from a given image block. They form a 50-dimensional feature vector. Documents in training and test set had no content overlap.

Reported results are averaged over 5 runs. Fig. 5 shows examples of text blocks for some fonts.

Fig. 5 Fonts used in experiments (From left to right – up to buttom): Badr, Divani Mazar, Jadid, Titr, Siavash Mazar, Nasim, Homa

TABLE I
Recognition Rate of WED Classifier on Typefaces Combined with Font Styles

	Regular	Italic	Bold	Bold Italic	Average
Badr	0.93	0.80	0.73	0.70	0.71
Divani Mazar	0.76	0.90	0.86	0.86	0.82
Jadid	0.60	0.36	0.50	0.63	0.46
Titr	0.33	0.66	0.33	0.36	0.41
Siavash Mazar	0.40	0.30	0.76	0.53	0.47
Nasim	0.76	0.63	0.63	0.66	0.68
Homa	0.90	0.80	0.80	0.83	0.80
Average	0.80	0.62	0.62	0.59	0.62

TABLE II
TYPEFACE CONFUSION MATRIX USING WED CLASSIFIER

	Badr	Divani Mazar	Jadid	Titr	Siavash Mazar	Nasim	Homa
Badr	.100	0	0	0	0	0	0
Divani Mazar	.04	.88	.08	0	0	0	0
Jadid	0	0	.79	.21	0	0	0
Titr	0	0	.04	.79	.05	.04	.08
Siavash Mazar	0	.04	.07	.12	.69	0	.08
Nasim	0	0	0	0	0	.100	0
Homa	0	0	0	0	0	.08	.92

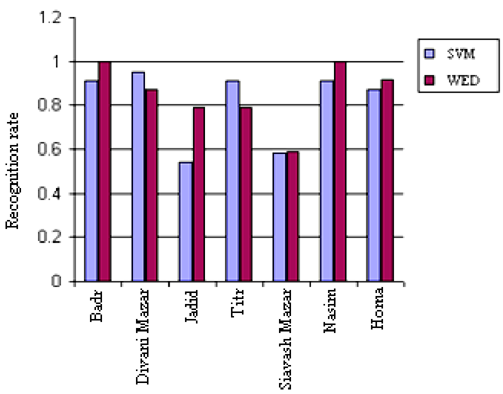

Fig. 6 Recognition rate of SVM classifier compared with WED over typefaces

IV. CONCLUSION AND FUTURE WORKS

Application of Gabor filters for Persian font recognition is considered in this paper. Global texture feature are used for recognition. The method is content independent, so the contents of text images are not necessary to have content overlap.

Because there is not much difference between different font styles for some fonts (Jaded, Titr, Siavash Mazar) recognition rate over those fonts are poor in comparison with other typefaces (As can be seen from results in Table I). Table II shows that Gabor filters are very capable to recognize the typeface of a font. Table III shows the superiority of WED classifier for font recognition task over SVM classifier.

As a future work one can consider combination of the global and local features. For example for typefaces with less difference over font-styles, a 2 step mechanism could be helpful. Gabor filters can recognize the typeface of a font in first step, and in the second step local features can determine the font style of that font.

REFERENCES

[1] B. B. Chaudhuri, and U. Garain, “Automatic Detection of Italic, Bold and All-Capital words in Document Images”, *Proc. of 14th International Conference on Pattern Recognition* - ICPR, Brisbane (Australia), pp.610-612, 1998.

[2] B. B. Chaudhuri, and U. Garain, “Extraction of type style-based meta-information from imaged documents”, *International Journal on Document Analysis and Recognition* - *IJDAR*, vol. 3, no. 3, pp. 138-149, 2001.

[3] K. Y. Wong, R. G. Casey, and F. M. Wahl, “Document Analysis System”, *IBM Journal of Research and Development,* vol. 26, no. 6, pp. 647-656, 1982.

[4] V. Eglin, S. Bres, and H. Emptoz, “Statistical characterization and classification of printed text in a multiscale context”, Proc. of 2nd *International Workshop on Statistical Techniques in Pattern Recognition* - SPR’98, Sydney (NSW), (Australia), pp.960-967, 1998.

[5] L. Duffy, “Recherche d'information logique dans les documents à typographie riche et récurrente. Application aux sommaires”, *Thesis of the Institut National des Sciences Appliquées* - INSA de Lyon, Lyon (France), 160p, 1997.

[6] A. Zramdini, “Study of Optical Font Recognition Based on Global Typographical Features”, *Thesis of the University of Fribourg, Fribourg (Switzerland)*, 170p, 1995.H. Poor, An Introduction to Signal Detection and Estimation. New York: Springer-Verlag, 1985, ch. 4.

[7] D. S. Doermann, E. Rivlin, and A. Rosenfeld, ”The function of documents”, International Journal of Computer Vision - IJCV, pp. 799-814, 1998.E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” IEEE Trans. Antennas Propagat., to be published.

[8] V.Wu, R. Manmatha, and E. M. Riseman, “Finding Text in Images”, *Proc. of Second ACM International Conference on Digital Libraries,* Philadelphia, PA (USA), pp.23-26, 1997.

[9] A. K. S. Jain, and K. Bhattacharjee, “Address Block Location on Envelopes Using Gabor Filters”, *Pattern Recognition*, vol. 25, no. 12, pp. 1459-1477, 1992.

[10] A. K. S. Jain, K. Bhattacharjee, and Y. Chen, “On texture in document images”, *Proc. Of IEEE Conference on Computer Vision and Pattern Recognition, Champaign*, Illinois (USA), pp.677-680, 1992.

[11] A. Bénédicte, H. Emptoz, “Font Type Extraction and Character Prototyping Using Gabor Filters”. *International Conference on Document Analysis and Recognition*, p, 799-803, 2003.

[12] Y. Zhu, T. Tan, and Y. WANG, “Font Recognition Based On Global Texture Analysis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1192-1200, 2001.

[13] N. Otsu, “A threshold selection method for gray-level histogram”, *IEEE Trans. On System Man Cybernetics*, vol. SMC-9, no. 1, pp. 62-66, 1979.

[14] N. Petkov, “Biologically motivated computationally intensive approaches to image pattern recognition”, *Future Gen. Comput. Syst.*, vol. 11, pp. 451-465, 1995.

[15] T. N. Tan, “Texture feature extraction via visual cortical channel modeling”, *Proc. Of 11th IAPR Int. Conf. Pattern Recognition*, vol. 3, pp. 607-610, 1992.

[16] V. Vapnik, “The Nature of Statistical Learning Theory”, Springer, 1995.

[17] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition”, Data Mining and Knowledge Discovery, vol. 2, no.2, pp. 1-47, 1998.

[18] V. Vapnik, “Statistical learning theory”. John Wiley and Sons, New York, 1998.

Ali Borji received the B.S. degree in computer engineering (software) from Petroleum University of Technology, Tehran, Iran, in 2001, and the M.S. degree in computer engineering (artificial intelligence) from Shiraz University, Shiraz, Iran in 2004. He is now working toward his Ph.D. at the School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran. His research interests include computational neuroscience, human and machine vision, machine learning, pattern recognition, and biologically motivated algorithms.

Mandana Hamidi received the B.S. degree in computer engineering (software) from Azad University of Shiraz, Shiraz, Iran, in 2000, and she is working toward her M.S. degree in computer engineering (artificial intelligence) in Azad University of Qazvin, Qazvin, Iran. Her research interests include human and machine vision, machine learning, pattern recognition, and biologically motivated algorithms.