# Off-line English and Chinese Signature Identification Using Foreground and Background Features

Srikanta Pal School of Information and Communication Technology, Griffith University, Queensland, Australia. Email: srikanta.pal@griffithuni.edu.au Umapada Pal Computer Vision and Pattern Recognition Unit, Indian Statistical Institute Kolkata-700108, India. Email: umapada@isical.ac.in Michael Blumenstein School of Information and Communication Technology, Griffith University, Queensland, Australia. Email: m.blumenstein@griffith.edu.au

Abstract— In the field of information security, the usage of biometrics is growing for user authentication. Automatic signature recognition and verification is one of the biometric techniques, which is only one of several used to verify the identity of individuals. In this paper, a foreground and background based technique is proposed for identification of scripts from bi-lingual (English/Roman and Chinese) off-line signatures. This system will identify whether a claimed signature belongs to the group of English signatures or Chinese signatures. The identification of signatures based on its script is a major contribution for multi-script signature verification. Two background information extraction techniques are used to produce the background components of the signature images. Gradient-based method was used to extract the features of the foreground as well as background components. Zernike Moment feature was also employed on signature samples. Support Vector Machine (SVM) is used as the classifier for signature identification in the proposed system. A database of 1120 (640 English+480 Chinese) signature samples were used for training and 560 (320 English+240 Chinese) signature samples were used for testing the proposed system. An encouraging identification accuracy of 97.70% was obtained using gradient feature from the experiment.

Keywords: Off-line verification systems, Signature identification, biometrics, authentication systems, SVM.

## I. INTRODUCTION

Signatures are a socially accepted authentication medium and they are widely used as proof of identity in our daily life. Automatic signature recognition by computers has received wide research interests in the field of pattern recognition. There are two different ways to recognize the signature: verification and identification. Verification involves confirming or denying a person's claimed signature. On the other hand, identification decides the signature group among the number of groups where the claimed signature belongs.

Today, biometric technologies are increasing and more commonly being used to ensure identity authentication of access to sensitive data. For historical reasons, the handwritten signature continues to be the most commonly accepted form of transaction confirmation, as well as being used in civil law contracts, acts of volition, or authenticating one's identity. Signature verification has been a topic of intensive research during the past several years [1] due to the important role it plays in numerous areas, including in the financial system.

Signatures have been accepted as an official means to verify personal identity for legal purposes on such documents as cheques, credit cards, wills etc. The handwritten signature is therefore well established and accepted as a behavioural biometric. Considering the large number of signatures verified daily through visual inspection by people, the construction of a robust and accurate automatic signature verification system has many potential benefits for ensuring authenticity of signatures and reducing fraud and other crimes.

The goal of a signature authentication system is to verify the identity of an individual based on an analysis of his or her signature through a process that discriminates a genuine signature from a forgery. The identification/verification of human signatures is particularly concerned with the improvement of the interface between human-beings and computers [2]. A signature identification/verification system and the associated techniques used to solve the inherent problems of authentication can be divided into two classes [3]: (a) on-line methods [4] to measure temporal and sequential data by utilizing intelligent algorithms [5] and (b) off-line methods [6] that use an optical scanner to obtain handwriting data written on paper. Off-line signature identification/verification deals with the identification/verification of signatures, which appear in a static format [7]. On-line signature identification/verification has been shown to achieve much higher rates than off-line one [6], as a considerable amount of dynamic information is lost in the off-line mode. But off-line systems have a significant advantage as they do not require access to special processing devices when the signatures are produced. Moreover, the off-line method has many more practical application areas than that of the on-line variety.

The use of signatures has been one of the more convenient methods for the identification and verification of human beings. Signatures represent a particular writing style and very often are a combination of symbols and strokes. So it is obviously necessary to deal with a signature as a complete image with a special distribution of pixels, representing a particular writing style and is not considered as a collection of letters and words [8]. It is often difficult for a human to instantly verify two signatures of the same person because signature samples from the same person are similar but not identical, and signatures can change depending on elements such as mood, fatigue and time. In addition, a person's signature often changes radically during their lifetime. Great inconsistency can even be observed in signatures according to country, habits, psychological or mental state, physical and practical conditions [9].

There are many pieces of work on script identification. K. Roy et al. [10] proposed a system for word-wise handwritten script identification for Indian Postal automation. Using matra/Shirorekha, water reservoir concept based feature, etc. a tree classifier was generated for Bangla/Devnagari and English word-wise scripts identification. Hochberg, et al. [11] proposed an algorithm for script and language identification from handwritten document images using statistical features based on connected component analysis. Hangarge and Dhandra [12] investigated a texture as a tool for determining the script of handwritten document image, based on the observation that text has a distinct visual texture. Further, K nearest neighbour algorithm was used to classify 300 text blocks as well as 400 text lines into one of the three major Indian scripts: English, Devnagari and Urdu, based on 13 spatial spread features extracted using morphological filters. Their proposed algorithm achieved average classification accuracy as high as 99.2% for bi-script and 88.6% for tri-script separation at text line and text block level respectively with five fold cross validation test. Roy and Pal [13] presented an automatic scheme for word-wise identification of handwritten Roman and Oriya scripts for Indian postal automation. In their proposed scheme, a piecewise projection method was used for line and word segmentation. Finally, using different features like, water reservoir concept based features, fractal dimension based features, topological features, scripts characteristics based features etc., a Neural Network (NN) classifier was used for word-wise script identification. For experiment, 2500 words were considered and overall accuracy of 97.69% was obtained from the proposed identification scheme. Although there are many pieces of work on script identification for general text, to the best of our knowledge there is no work of script identification for signature written in Chinese and English scripts.

Numerous techniques for feature extraction and classification have been put forward in the literature for the processing of signatures. Justino et al. [14] proposed an offline signature verification system based on Hidden Markov Models (HMMs) to detect random, casual, and skilled forgeries. Three features: a pixel density feature, a pixel distribution feature and an axial slant feature are extracted from a grid segmentation scheme. Lv *et al.* [15] developed a Chinese off-line signatures. Support Vector Machines were employed for classification. Four different types of features such as Moment feature, Direction feature, Gray distribution and Stroke width distribution feature were used in their research. Nguyen *et al.* [16] developed an off-line signature verification system based on global features and the Support Vector Machine classifier. In their paper, the combination of the Modified Direction Feature (MDF) with three global features: a feature using Energy information, maxima feature, and ratio feature are reported. In addition, the survey by Weiping et al. [17] summarises some additional features and approaches that have been previously investigated.

The remainder of this paper is structured as follows. Significance of multi-script signatures is described in Section II. Section III deals with describing the signature database. Section IV discusses the feature extraction technique employed. Section V introduces the experimental settings. Details of the classifier used are presented in Section VI. Results and discussions are given in Section VII and error analysis is detailed in Section VIII. Finally, conclusions and future work are discussed in Section IX.

# II. SIGNIFICANCE OF MULTI-SCRIPT SIGNATURE IDENTIFICATION

Although, many systems involving off-line signature recognition and verification have been developed, all of these systems have solely considered single-script signatures. However, signatures may be written in different languages and there is a need to undertake a systematic study in this area. In the field of signature verification, most of the published work has been undertaken for English signatures. Only a few studies have been performed for Chinese, Japanese, Persian and Arabic signatures [18-21]. As indicated earlier, researchers have used different features for signature verification and it was noted that all the published work is based on foreground information.

Moreover, to the best of the authors' knowledge, there is no published work employing signatures written in two different languages. However, sometimes the signatures of different scripts are desired for official transactions. Some countries have more than one or two scripts that are not only used for handwriting but also for signing purposes. A multilingual country like India has many different scripts such as such as Hindi, Bangla, Telugu, Tamil etc. that are used for writing as well as for signing purposes based on different locations or regions of the country. So, there is a need to work on signatures written in two or more languages, especially in off-line signature identification and verification, considering signatures of two or more scripts. In this paper, a signature identification system is proposed for two scripts: English and Chinese. To the authors' knowledge, background information has not so far been used in signature identification research, and this proposed system is one of the first of its kind that uses background and foreground information. Some English and Chinese signature samples are shown in Fig. 1 and Fig. 2, respectively.

## III. SIGNATURE DATABASE

Although automatic signature verification has been an active research area for several decades, there is a lack of commonly used, publicly-available signature databases for signature identification and verification systems. The signatures of English and Chinese scripts are considered for this signature identification approach. For experiment, it was necessary to create a custom database, which included English as well as Chinese signatures.

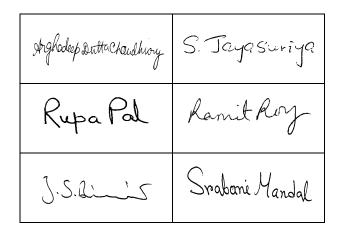


Figure 1. Examples of Some English Signatures

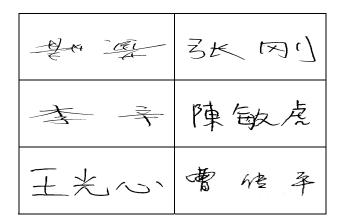


Figure 2. Examples of Some Chinese Signature

# A. Pre-processing

The signatures to be processed by the system should be in a digital image format. At the very beginning, the images were captured in 256 level grey-scale at 300 dpi and stored in TIFF format (Tagged Image File Format) for the purpose of future processing. In the pre-processing step, a histogram-based threshold technique is applied for binarization. In this step, the digitized grey-scale images are converted to a two-tone image. Then the signature images are extracted from

the document forms used to collect the signatures. A signature collecting form with signatures in gray level is shown in Fig. 3. Each extracted binary signature image is also stored in TIFF format. A typical scanned and binarized signature is shown in the Fig. 4 and Fig. 5 respectively.

3长 网)	张顺	张凤
张冈山	张网	3长 冈!
张刚	张网	张冈门
张刚	张冈	3代 (月1)
张刚	张冈	张图
张冈山	张冈山	张网
张网	张刚	张网
张冈山	张冈山	张网

Figure 3. Signature Collecting Form with Signatures



Figure 4. Example of a Scanned Signature Image



Figure 5. Example of Binary Image of the Signature shown in Fig. 4.

#### B. Data collection and database preparation

For signature identification of English and Chinese scripts 960 English signatures was collected from 50 individuals from different parts in West Bengal, India. Another dataset of 720 Chinese signatures was collected from 20 individuals from Australia. The training and test samples were allocated as nearly 66.66% and 33.33% respectively of the total signature samples for each script.

#### **IV. FEATURE EXTRACTION**

Feature extraction is an important part in any pattern recognition system. The choice of a powerful set of features is crucial step in signature verification systems. In this work, Gradient features and Zernike Moment features are extracted from the signature images for the purpose of signature identification.

### A. 400 Dimensional Gradient Feature

The gray-scale local-orientation histogram of the component is used for 400 dimensional feature extractions. These features help us to compute segmentation confidence. To obtain 400 dimensional features the following steps are employed.

Step 1: At first size normalization of the input binary image is done. Here we normalize the image into  $126 \times 126$  pixels.

Step 2: The input binary image is then converted into a grayscale image applying a  $2 \times 2$  mean filtering 5 times.

Step 3: The gray-scale image is normalized so that the mean gray scale becomes zero with maximum value 1.

Step 4: Normalized image is then segmented into  $9 \times 9$  blocks.

Step 5: A Roberts filter is then applied on the image to obtain gradient image. The arc tangent of the gradient (strength of gradient) is quantized into 16 directions (an interval of  $22.5^{\circ}$ ) and the strength of the gradient is accumulated with each of the quantized direction. By strength of Gradient (f(x,y)) we mean

$$f(x,y) = \sqrt{(\Delta u)^2 + (\Delta v)^2}$$

and by direction of gradient  $(\theta(x, y))$  we mean  $\theta(x, y) = tan^{-1} \frac{\Delta v}{\Delta u}$ , here

$$\Delta u = g(x + 1, y + 1) - g(x, y),$$
  

$$\Delta v = g(x + 1, y) - g(x, y + 1)$$

and g(x,y) is a gray scale value at an (x,y) point.

Step 6: Histograms of the values of 16 quantized directions are computed in each of  $9 \times 9$  blocks.

Step 7:  $9 \times 9$  blocks is down sampled into  $5 \times 5$  by a Gaussian filter. Thus, we get  $5 \times 5 \times 16 = 400$  dimensional feature.

The Zernike Moment and Gradient feature extractions are separately employed on foreground parts to get two feature sets. Two different background parts of signature images (discussed as following steps) are used in Gradient feature extraction method to get another two feature sets.

### Step 1. Foreground information

In this step, Gradient features and Zernike Moment features are extracted from the foreground part of both English and Chinese signature images. After scanning a document, the grey image is binarized to obtain the binary data for further processing. The black pixel portion of the binary image within the bounding box is referred to as the foreground part.

## **Step 2. Background Information**

In this step, the Gradient features based on the background part of the signatures are extracted.

To obtain the background part, all the columns of the input image are scanned from top to bottom and from bottom to top to get the uppermost and lowermost black pixels. The portion between the uppermost and lowermost black pixels is converted into black. In this way the background portion of every signature is detected. This converted portion containing black pixels is called the background area. This background part of the images obtained in step 2 is called BG<sub>TB.</sub> (Background part from top to bottom and from bottom to top scanning). Such background part obtained from original signatures (Fig.6 and Fig.8) is shown in Fig.7 and Fig.9 for two different signatures.

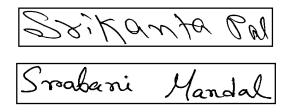


Figure 6. Original English Signature samples

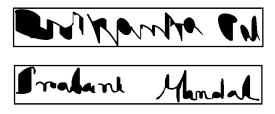


Figure 7. Background Part of English Signature (derived from fig.6.)



Figure 8. Original Chinese Signature samples





Figure 9. Background Part of Chinese Signature (derived from fig.8.)

### **Step 3: Extended Background Information**

In this step, the Gradient features based on the extended background part of the signatures are extracted.

For this extended background part extraction, a different technique is applied. A subtraction operation is performed on two different background parts of the signature images to get this information and we call it as extended background.

Here, all the columns of the input image are scanned from top to bottom and from bottom to top to get the uppermost and lowermost black pixels. The portion between the uppermost and lowermost black pixels is converted into black. All the rows of this resultant image are scanned from left to right and from right to left to get the leftmost and rightmost black pixels. The portion between the leftmost and rightmost black pixels is also converted into black. In this way the background portion of every signature is detected. The resultant background part of the images obtained in this two step is called BG<sub>TBLR</sub>. (Background part from top to bottom, bottom to top, left to right and right to left scanning).

Then, the subtraction operation  $(BG_{TBLR}-BG_{TB})$  is performed in between the background part of  $BG_{TBLR}$ images and the background part of  $BG_{TB}$  images. The desired extended background parts of two signature images are shown in Fig.10 and Fig.11.



Figure 10. Extended Background Part of English Image



Figure 11. Extended Background Part of Chinese Image

Gradient features computed from foreground part, background part and extended background part are called here feature set-1, feature set-2 and feature set-3, respectively. Zernike Moment feature computed from foreground part are called here feature set-4. Detailed about Zernike features are given below.

#### **B.** Zernike Moments Feature

Zernike polynomials are an orthogonal set of complex-valued polynomials:

$$V_{nm}(x, y) = R_{nm}(x, y) \cdot \exp((jm \tan^{-1}(\frac{y}{x})))$$

where,

$$x^{2} + y^{2} \le 1, \ j = \sqrt{-1}, \qquad n \ge 0, \qquad |m| \le n$$

and n - |m| is even and Radial polynomials  $\{R_{nm}\}$  are defined as:

$$R_{nm}(x,y) = \sum_{s=0}^{\frac{n-|m|}{2}} B_{n|m|s}(x^2 + y^2)^{\frac{n}{2}-s}$$

where,

$$B_{n|m|s} = \frac{(-1)^s (S-1)!}{S! \left(\frac{n+|m|}{2} - S\right)! \left(\frac{n-|m|}{2} - S\right)!}$$

The complex Zernike moments of order n and repetition m are given by:

$$A_{nm} = \frac{n+1}{\pi} \sum_{x} \sum_{y} f(x, y) V_{nm}^*(x, y)$$

Where  $x^2 + y^2 \le 1$  and symbol \* denotes the complex conjugate operator [24].

The Zernike moments can be computed by the scale invariant Central moments as follows:

$$A_{nm} = \frac{n+1}{\pi} \sum_{\substack{k-|m|\\n-k=even}}^{n} \sum_{a=0}^{b} \sum_{d=0}^{|m|} (-j)^{d}$$
$$\cdot \left( \binom{m}{d} \right) \binom{b}{a} B_{n|m|s} G_{k-2a-d,2a+d}$$

where,

$$b = \frac{n - |m|}{2} - S \quad and \quad j = \sqrt{-1}$$

# V. EXPERIMENTAL SETTINGS

Here experimental part of signature verification technique consists of four settings. The number of signature samples used for training and testing purposes in each setting of the experiment is described in Table 1.

TABLE 1. NO. OF SAMPLES IN TRAINING AND TESTING PHASE

Settings	Phases	English Signature Samples	Chinese Signature Samples
Cattinga 1	Training	640	480
Settings 1	Testing	320	240
Settings 2	Training	640	480
	Testing	320	240
Sattings 2	Training	640	480
Settings 3	Testing	320	240
Sottings 1	Training	640	480
Settings 4	Testing	320	240

# VI. CLASSIFIERS DETAILS

## SVM Classifier

In our experiments, we have used a Support Vector Machine (SVM) classifier. The SVM is originally defined for two-class problems and it looks for the optimal hyper plane which maximizes the distance and the margin, between the nearest examples of both classes, named support vectors (SVs). Given a training database of M data:  $\{x_m | m=1,..., M\}$ , the linear SVM classifier is then defined as:

$$f(x) = \sum_{j} \alpha_{j} x_{j} \cdot x + b$$

where  $\{x_j\}$  are the set of support vectors and the parameters  $\alpha_j$  and *b* have been determined by solving a quadratic problem [22]. The linear SVM can be extended to various non-linear variants; details can be found in [22, 23]. In our experiments, the Gaussian kernel SVM outperformed other non-linear SVM kernels, hence we are reporting our recognition results based on the Gaussian kernel only.

#### VII. RESULTS AND DISCUSSION

An investigation of the performance of a signature identification system involving English and Chinese off-line signatures are presented here. The Gradient feature, Zernike moment feature and SVM classifiers were employed and encouraging results were obtained. Table 2 shows the results obtained in this expriment from four different feature sets. From Table 2 it can be noted that the best result(97.70%) was obtained using feature set 2.

As no research had been done based on bi-script signature identification, it is difficult to compare the performance of this identification system.

TABLE 2. ACCURACIES BASED ON DIFFERENT FEATURE SETS.

	<b>Recognition rates (%)</b>			
Classifier	Feature Set 1	Feature Set 2	Feature Set 3	Feature Set 4
SVMs	97.50	97.70	96.76	96.07

#### VIII. ERROR ANALYSIS

Confusion matrices obtained using the SVM classifier based on different feature sets are shown in Table 3 to Table 6. It is noted that only 11 Chinese signatures were misidentified as English signatures when feature set 2 is used. A sample of errors obtained for feature set 1 is shown in Fig.12. From Fig.12, it is evident that some of the components of the Chinese signature images look like English characters, and this caused misidentification of those signatures. We noted from this experiment that most of the errors occurred because of misclassification of Chinese signatures as English signatures.

TABLE 3. CONFUSION MATRIX OF FEATURE SET 1

	English	Chinese
English	318	2
Chinese	14	226

TABLE 4. CONFUSION MATRIX OF FEATURE SET 2

	English	Chinese
English	320	0
Chinese	11	229

TABLE 5. CONFUSION MATRIX OF FEATURE SET 3

	English	Chinese
English	320	0
Chinese	12	228

TABLE 6. CONFUSION MATRIX OF FEATURE SET 4

	English	Chinese
English	310	10
Chinese	12	228

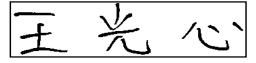


Figure 12. An Error Sample Obtained from the Classification.

## IX. CONCLUSIONS AND FUTURE WORK

This paper presents an off-line signature identification scheme of bi-script signatures. Gradient feature extraction from foreground and different background parts of signature images and the SVM classifier were utilized for this off-line signature identification scheme. Zernike Moment feature was also employed for foreground parts of signatures images. Encouraging results of 97.50% and 96.07% were obtained using foreground part from Gradient and Zernike Moment features respectively. The accuracies of 97.70% and 96.76% were obtained using two different background feature sets utilizing Gradient features. To the best of the authors' knowledge, such background features have not been used for the task of signature verification/identification and this is the first work using background features in this area. This scheme of bi-script off-line signature identification is also a novel contribution to the field of signature verification. In the near future, we plan to extend our work for multi-script offline signature verification.

#### REFERENCES

- S. Chen, and S. Srihari, "Use of Exterior Contour and Shape Features in Off-line Signature Verification", 8th ICDAR, pp. 1280-1284, 2005.
- [2] M.A. Ferrer, J.B. Alonso and C. M. Travieso, "Off-line Geometric Parameters for Automatic SignatureVerification Using Fixed-Point Arithmetic" IEEE PAMI, Vol.27, no.6, pp. 993-997, 2005.
- [3] S. Madabusi, V. Srinivas, S. Bhaskaran and M. Balasubramanian, "On-line and off-line signature verification using relative slope algorithm", International Workshop on Measurement Systems for Homeland Security, pp. 11-15, 2005.
- [4] S. Emerich, E. Lupu and C. Rusu. "On-line Signature Recognition Approach Based on Wavelets and Support Vector Machines", International Conf. on Automation Quality and Testing Robotics, pp.1-4, 2010.
- [5] A. Kholmatov, and B. Yanikoglu, "Identity Authentication using improved online signature verification method", Pattern Recognition Letters, Vol-26, pp 2400-2408, 2005.
- [6] M. Kalera, S. Srihari, and A. Xu. "Offline signature verification and identification using distance statistics", IJPRAI, pp.1339-1360, 2004.
- [7] I. Pottier and G. Burel, "Identification and Authentication of Handwritten Signatures with a Connectionist Approach", In Proc. IEEE Conf. on Neural Networks, pp. 2948-2951, 1994.
- [8] B. Fang, C.H. Leung, Y.Y. Tang, K.W. Tse, P.C.K. Kwok and Y.K. Wong, "Off-line signature verification by the tracking of feature and stroke positions", Pattern Recognition, pp. 91-101,vol-36, 2003.
- [9] Katrin Franke, "Analysis of Authentic Signatures and Forgeries", In Proc. IWCF, pp 150-164, 2009.
- [10] K. Roy, A. Banerjee and U. Pal, "A System for Wordwise Handwritten Script Identification for Indian Postal Automation", In Proc. IEEE India Annual Conference (INDICON-04), pp. 266-271, 2004.

- [11] Judith Hochberg, Kevin Bowers, Michael Cannon and Patrick Keely, "Script and language identification for handwritten document images," IJDAR, vol.2, pp. 45-52, 1999.
- [12] Mallikarjun Hangarge and B. V. Dhandra, "Offline Handwritten Script Identification in Document Images", International Journal of Computer Applications, Volume 4 No.6, pp. 6-10, 2010.
- [13] K. Roy U. Pal, "Word-wise Hand-written Script Separation for Indian Postal automation", In Proc. 10<sup>th</sup> International Workshop on Frontiers in Handwritten Recognition (IWFHR), pp. 521-526, 2006.
- [14] E. Justino, E. Bortolozzi, R Sabourin, "Off-line Signature Verification Using HMM for Random, Simple and Skilled Forgeries," Proceedings of 7<sup>th</sup> ICDAR, pp. 105-110, 2001.
- [15] H. Lv, W. Wang, C. Wang, and Q. Zhuo, "Off-line chinese signature verification based on support vector machines," Pattern Recognition Letters, vol. 26, no. 15, pp. 2390–2399, 2005.
- [16] V. Nguyen, M. Blumenstein, G. Leedham, "Global Features for the Off-Line Signature Verification Problem", Proceedings of 10<sup>th</sup> ICDAR, pp.1300-1304, 2009.
- [17] H. Weiping, Y. Xiufen, and W. Kejun, "A survey of off-line signature verification," Proceedings of the Intl. Conf. on Intelligent Mechatronics and Automation, pp. 536-541, 2004.
- [18] Jun Lin and Jie-gu Li, "Off-Line Chinese Signature Verification", Int. Conference on Image Processing, pp. 205 – 207, 1996..
- [19] I.Yoshimura and M.Yoshimura, "Off-line verification of Japanese signatures after elimination of background patterns", Progress In Automatic Signature Verification, pp 53-68, 1993.
- [20] S. Ghandali and M.E. Moghaddam, "A Method for Off-line Persian Signature Identification and Verification Using DWT and Image Fusion", pp. 315 – 319. 2008.
- [21] M.A. Ismail and Samia Gad, "Off-line Arabic Signature Recognition and Verification", Pattern Recognition vol. 33, pp.1727-1740, 2000.
- [22] V.Vapnik, "The Nature of Statistical Learning Theory", Springer Verlang, 1995.
- [23] C. Burges, "A Tutorial on support Vector machines for pattern recognition", Data mining and knowledge discovery, pp.1-43, 1998.
- [24] C. H. Teh and R. T. Chin, "On Image Analysis By The Methods Of Moments," IEEE Trans. Patt. Anal. Mach. Intell., Vol. 10, No. 4, pp. 496-513, 1988.