USING STATISTICAL MOMENTS AS INVARIANTS FOR EYE DETECTION

Saideh Ferdowsi and Alireaz Ahmadyfard

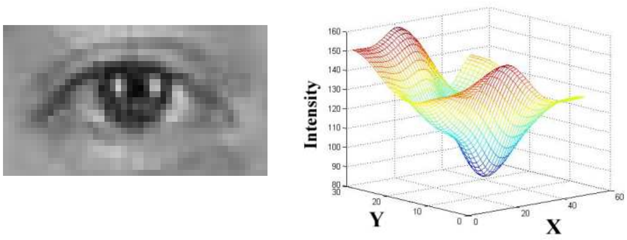
Department of Electrical Eng. Shahrood University of Technology
Shahrood,Iran
saideh_ferdosi@yahoo.com, ahmadyfard@shahroodut.ac.ir

ABSTRACT

In this paper we address the problem of eye detection in greyscale images. We represent face image using topographic labels to alleviate detection under severe lighting condition. The regions in topographic image are then described using regional invariant moments. The employed moments are invariant to similarity transform. This enables the proposed eye detection method to work under head movement. In detection phase we first provide a candidate list of points with pit label in topographic image. Image at neighbourhood of each pair of pit points are compared with eyes model using their corresponding feature vectors. Using a Bayesian classifier we detect the pair of points with the descriptors most similarity to the eyes. The result of experiments confirms the capability of proposed method for detecting eyes in face images.

1. INTRODUCTION

Eye detection in 2D images is a crucial step in many machine vision applications such as face detection, face recognition, expression analysis. A proper solution for the addressed problem plays an important role in developing systems for new applications such as measuring awareness of car drivers, human computer interaction (HCI), video conferencing and disabled people aiding system. The success of these applications directly depends on accuracy and robustness of eye detection. Significant variation of eye appearance in image which is result of eye size, position of head, eye closing, lighting condition and occlusion by hair and frame of glasses, makes eye detection a challenge. The methods proposed for eye detection in literature are classified in three categories: template based, feature based and appearance based methods. In the first group of methods a template based on sketch of eye model is constructed. A given face image is matched against the template to find the eyes location [1]. The success of this approach totally depends on consistency of the model and eyes from lighting condition, rotation and scaling points of view. Regardless of this problem, computational complexity of this approach is high. The attempts in feature based methods are to search for discriminative eye features in the image such as eye corners, intensity of iris or colour distribution of iris [2]. These methods fail in case of partially occlusion or rotation of head in depth. Finally an appearance based eye detection method aims to learn eye image using raw images. The learnt system is then ready to search for presence of eye in image [3-5]. This approach requires a significant number of learning images to handle variations such as rotation, size and scaling for eye detection.



(a) (b)

Figure 1. (a) intensity image of eye (b) eye representation as a surface in 3D space

In this paper we propose a method similar to the last group which learns eye model using its examples. But instead of using intensity image to learn eye model we use topographic features. In this representation an intensity image is considered as 3D function as shown in Figure 1.

As seen from this figure eye surface has an obvious local minimum (pit) at centre surrounded by hillside features.

This gives us a hint that eyes can be detected by exploring their terrain features.

Wang et al [6] proposed a method for eye detection based on terrain feature matching. In this method, first a face image is represented using topographic features from which a terrain map is generated. The terrain map is composed of topographic labels. Second, similarity between the terrain map of eye model and that of the test image is measured. The authors simply use the statistical distribution of terrain features for comparing two terrain maps.

There are some problems with the way of comparing the topographic features in Wang et al method. First, the distribution of terrain features do not characterise shape and geometrical distribution of a topographic label. So many regions in image may have similar statistical distribution for terrain features. Second the statistical distribution of terrain features is not invariant to geometrical transformations. So simply scale change between eye model and face image degrades the eye detection performance.

In this paper we propose a matching algorithm for topographic features which is invariant to similarity transform. The geometrical invariance simply can be extended to affine transform. We extract statistical features from each labelled region in topographic map. Using the collection of extracted features for all labels in topographic map we construct a geometrically invariant features vector which describes image. In detection phase we compare the feature vector extracted from eye model with that of the face image to detect eyes.

The paper is organized as follows. In the next section we explain methodology for eye detection. In Section 3 we present the result of experiments when the proposed method is applied on face image. In this section we also compare the proposed method with Wang et al [6] method. Finally we draw the paper to conclusion in Section4.

2. METHODOLOGY

In this section we explain our method for representing an image using topographic labels. We propose a method to describe the labelled image using a feature vector which is constructed from statistical invariant moments. In order to detect eyes in the image we extract the pit points in the labelled image as candidates for centre of eye pattern. Then using a Bayesian classifier we measure possibility that each pair of candidate points being the centre of right and left eyes in the image. The distance between feature vectors extracted from the models and test image is used for measuring the similarity between patterns. In the next subsections we explain the proposed method in details.

2.1 Image representation using topographic labels

Using topographic models for representing images has been reported in computer vision literature [7-9]. This method is classified in appearance based category. The main advantage of using topographic features for representation respect to intensity is its robustness to lighting condition [9]. With change in lighting condition such as lighting source direction and its power the appearance of eyes in an intensity image totally changes. That is why an appearance based method need many examples to model an object.

Consider a grey scale face image as a surface in 3D space where *x* and *y* axes are along image dimensions. The value of surface at pixel (*x,y*) is the pixel intensity *f(x,y)*. Depending on the topographic property of surface at each pixel one of twelve topographic labels in Figure 2 is assigned to the pixel [7]. In order to label intensity image based on topographic property let us consider the input image as a continues function *f(x,y)*. The topographic label at each pixel of image is determined using first and second order derivatives on surface *f*.

Considering the Hessian matrix of this function as follows:

$$H(x,y) = \begin{bmatrix} \frac{\partial^2 f(x,y)}{\partial x^2} & \frac{\partial^2 f(x,y)}{\partial x \partial y} \\ \frac{\partial^2 f(x,y)}{\partial x \partial y} & \frac{\partial^2 f(x,y)}{\partial y^2} \end{bmatrix} \tag{1}$$

After applying eignvalue decomposition to the Hessian matrix we have:

$$H = UDU^T = [u_1 \quad u_2].diag(\lambda_1 \quad \lambda_2).[u_1 \quad u_2]^T \tag{2}$$

Where λ_1, λ_2 are the eigenvalues and u_1, u_2 are the orthogonal eigenvectors.

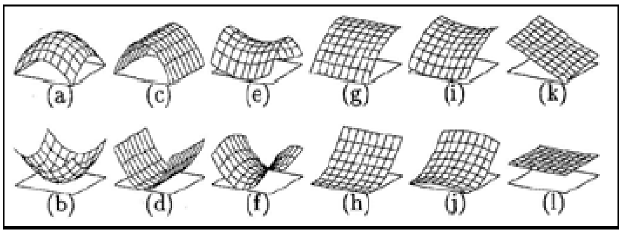


Figure 2. Topographic labels (a) peak (b) pit (c) ridge (d) ravine(e) ridge saddle (f) ravine saddle (g) convex hill (h) concave hill (i) convex saddle hill (j) concave saddle hill (k) slop hill and (l) flat

Using the eigenvalues, eigenvectors of the Hessian matrix and the vector of surface gradient $\nabla f(x,y)$ at each pixel the topographic label at each pixel is determined. For instance a pixel in the image takes the *pit* label if $\|\nabla f(x,y)\| = 0$ and both eigenvalues being positive. The similar conditions for gradient but negative values for the eigenvalues of the Hessian matrix indicate a *peak* label.

For further details on determining other topographic labels one can refer to Ref [9]. It is worth to note that image noise can cause an undesired result of topographic labelling. As shown by Wang et al [6] a smoothing filter before topographic labelling provide more acceptable result. For this purpose we filter input image using a Gaussian kernel before the labelling. The filter parameters should be selected based on the size of interest pattern (eye) and the level of input noise. However an extra-smoothing of intensity image causes missing some topographic features of the interest pattern.

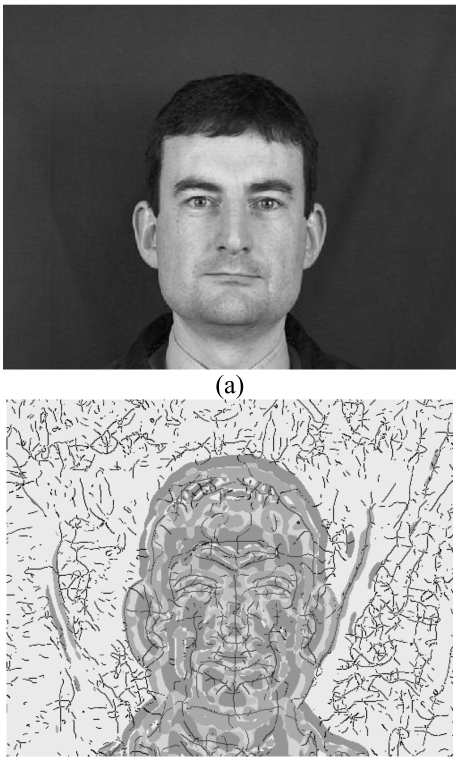
The above procedure is defined for continues surfaces; for digital images the continues derivatives of *f* must be estimated. We use a smoothed differentiation filter based on the Chebyshev polynomials to estimate derivatives of surface [10].

In this regard the *p* th and *q* th order derivative of *f* respect to variables *x* and *y* respectively is estimated using the following formula:

$$f^{(p,q)}(x,y) = \sum_{i=-N}^{N} \sum_{j=-N}^{N} f(x-i, y-j)h(i,p)h(j,q) \tag{3}$$

Where *f(x,y)* is the digital input image. Filters *h(i,p)* and *h(j,q)* are kernels for estimation of function derivatives along *x* and *y* directions respectivly using Chebyshev polynomials [10].

Using the above strategy we label input image based on topographic labelling. Figure 3, 4 show the face and eye image with the corresponding terrain map obtained from the topographic labelling. In order to detect eyes in face image we first extract *pit* pixels in the labelled image. Each pit pixel is a candidate for centre of eye. In order to find the place of true eyes among the pit candidates in face image we describe image at each pit neighbourhood using a feature vector and match with that of right and left eyes.


(a)

(b)

Figure 3. (a) intensity image of face and (b) terrain map obtained based on topographic labels

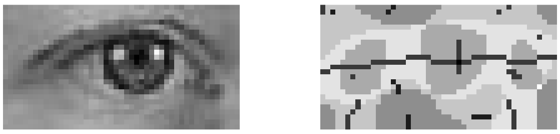

(a) (b)

Figure 4. (a) intensity image of eye and (b) terrain map obtained based on topographic labels

Wang et al [6] define a rectangular patch centred at each pit pixel and describe the image patch using statistical distribution of labels in terrain map. Our preliminary experiments showed that the extracted features are not discriminative enough; so many false positives are resulted.

In this paper we describe the image patch at each pit using invariant moments of regions in terrain map. The shape of each labelled region in terrain map is described using two invariant moments. Hu [11] introduced a set of features for describing a region. These features are proven to be invariant to similarity transformation. These features are defined based on geometrical moments of region. Let us denote corresponding region to l th topographic label in image patch byR_l. The *(p+q) th* geometrical moment this region is defined as follows:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p \, y^q R_1(x,y) \qquad p,q = 0,1,2,\ldots$$

$$\overline{x} = \frac{m_{10}}{m_{00}} \qquad and \qquad \overline{y} = \frac{m_{01}}{m_{00}} \tag{4}$$

In this regard the central moments of the region is given:

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} (x-\overline{x})^p (y-\overline{y})^q R_l(x,y) \tag{5}$$

From these moments Hu [11] extracted seven features which are invariant to similarity transform. The features are defined as follows :

$$\begin{aligned}
\phi_1 &= \mu_{20} + \mu_{02} \\
\phi_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
\phi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
\phi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
\phi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - \\
&\quad 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \\
&\quad \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\
\phi_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + \\
&\quad 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
\phi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - \\
&\quad 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \\
&\quad \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]
\end{aligned} \tag{6}$$

Through experiments we found that five features among twelve topographic features are more discriminative for eye regions. Hence we describe an eye image using these labels: ravine, convex hill, concave hill, convex saddle hill and concave saddle hill (Figure 2). Using the above features we construct a vector for an eye image including 35 features denoted by $\boldsymbol{V}$ (seven features for each of five labelled regions).

2.2 Eye detection

In training phase we collect three set of images corresponding to images of right eye, images of left eye and non-eye images. The feature vectors extracted from images in each training class is used to obtain a statistical model for the class. We assume that distribution of training samples in each class is a multimodal Gaussian function. Hence we have three normal distributions $(\Sigma_{R\_eye}, m_{R\_eye})$, $\mathcal{N}(\Sigma_{L\_eye}, m_{L\_eye})$ and $\mathcal{N}(\Sigma_{N\_eye}, m_{N\_eye})$ corresponding to right eye, left eye and non-eye classes. Using the training image in each class covariance matrix Σ and mean vector m is estimated.

For a given face image after extracting topographic labels we consider pixels with pit label as candidates for centre of eye. For each pair of pit pixels with distance d we define two rectangular patches centred at pit pixels with size *0.4d×0.6d.*

Images inside these patches are candidates for left and right eyes. The size of examined patch is selected based on the relative size of eyes and the distance between them. We describe the image in each patch using the defined feature vector. So a pair of vectors defines the left and right patch in a candidate pair $(\boldsymbol{V_L}, \boldsymbol{V_R})$.

Eye detection procedure is performed in two steps. First we classify each image patch of a candidate pair into one of three classes (left, right eyes and non-eye) using a Bayesian classifier. In other words, the Mahalanobis distance between a test feature vector $\boldsymbol{V_i}$ and the mean of each class is determined. The Mahalanobis distance between patch i (i=L or R) and class j (j=L_eye, R_eye, N_eye) is defined as follows:

$$D_{i,j}{}^2 = \left(V_i - m_j\right)^T \Sigma_j^{-1} \left(V_i - m_j\right) \qquad (7)$$

The input patch takes label of the class with minimum distance from the patch.

For a pair of candidates as eyes we expect the left patch and right patch being take the L_eye label and the R_eye label respectively. As in a face image more than one pair of candidate patches may pass this stage, we select only the one with minimum distance to the eyes model as the eyes. The distance of a pair of patches to the eyes model is defined as:

$$D_{pair} = \sqrt{D^2{}_{L,L\_eye} + D^2{}_{R,R\_eye}} \qquad (8)$$

3. EXPERIMENTS

We set an experiment to evaluate the performance of the method proposed for eye detection. We used the 1300 face images from XM2VTS database to test our method. For training of the classifier we used 400 images of right and left eyes and 600 images of non-eyes. The training faces are different from test set. The database consists of colour images of faces which we convert them to greyscale images in our experiment.

We provided the training images in our dataset by cutting manually eye and non-eye regions from the face images. The training set is grouped into left eye, right eye and non-eye classes. To selected training images in non-eye class we first label pixels in face images using topographic property. The pixels with pit label which are out of eye regions are detected. We extracted a rectangular patch centred at each detected pit pixel from the above procedure. The parameters for features in three clusters are estimated (Σ_j, m_j) $j =$ {L_eye,R_eye, N_eye }.

We tested the proposed eye detection algorithm on 1300 face images. In this experiment the proposed method based on regional moments is compared to Wang et al [6] method. Table 1 shows the result of eye detection for two methods.

As seen from the table detection rate for the proposed method 8% higher than this rate for Wang et al [6] method.

A number of test images (some are not in XM2VTS database) in and the result of eye detection using the proposed method are shown in Figure 5. This figure shows the results of eye detection under partial occlusion (a),(b) (by hair or classes) , with head rotation (c),(d) and sever lighting condition(e) and closed eyes (f). As seen the method can successfully cope with lighting change and the head rotation. We need to emphasis that the invariance property of topographic label to lighting condition makes the method robust to illumination change. On the other hands using invariant features to similarity transform enables the method to perform well under head rotation.

Table 1. The performance of two eye detection methods

Method	Detection rate	Fail to detect one eye	Fail to detect both eyes
Wang's method	82%	12%	6%
The proposed method	90%	7%	3%

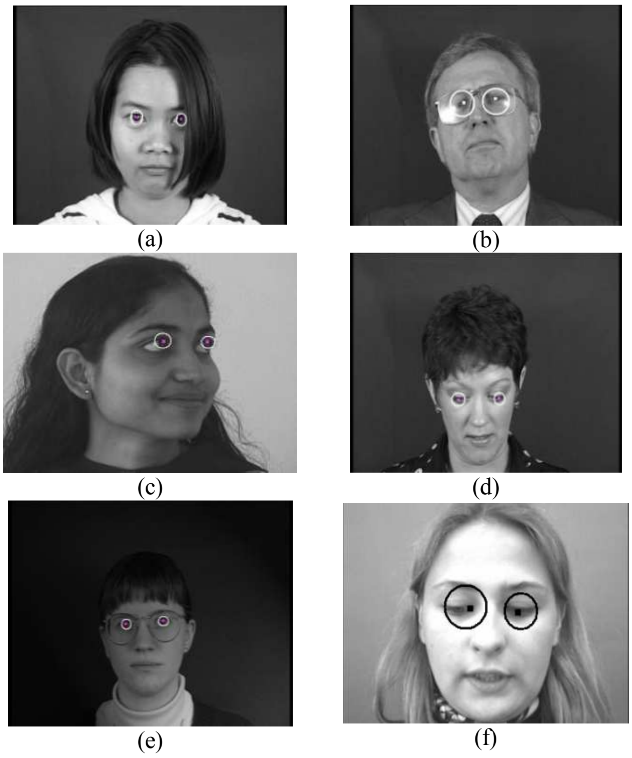
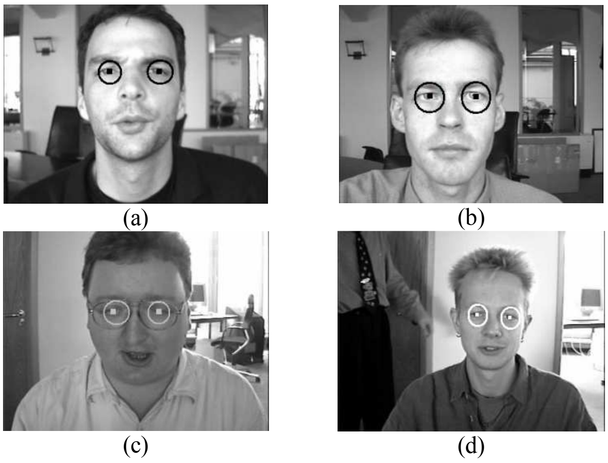


Figure 5. Some face images and the detected positions as eyes

We also evaluated robustness of the proposed method in face images taken in real scenes. The promising results confirmed that our method is also successful in images with background. Figure 6 shows some sample results of eye detection in images of this type. As seen from this figure, the position of eyes has been successfully detected in different scenes.



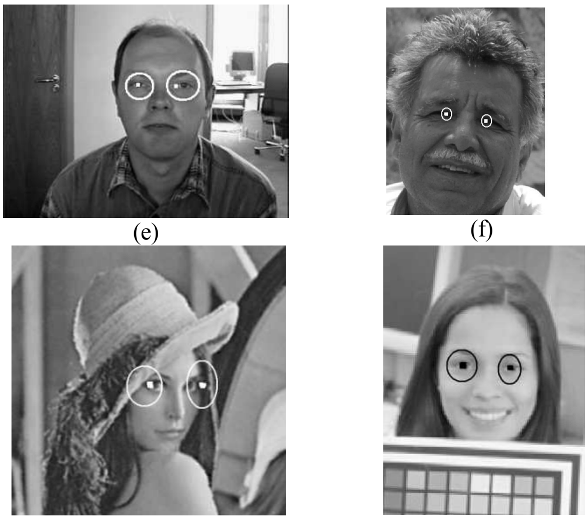
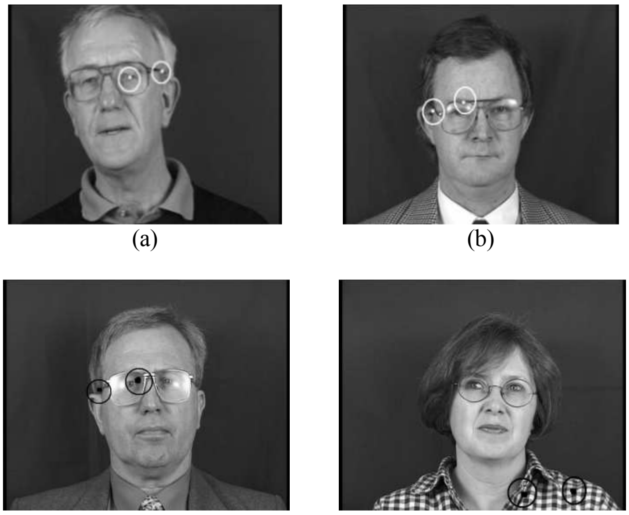

(e) (f)

Figure 6. Sample face images with real background and corresponding detected positions as eyes

Although the proposed method could successfully detect the position of eyes in variety of face images, there were also some failures. Figure 7 shows two examples for which the proposed eye detector fails. The main cause of fail is the reflection of classes and occlusion. Moreover, existing eye-like and textured backgrounds such as chessboard shirts, text, etc. leads to false detection in some cases.


(a) (b)

(c) (d)

Figure 7. Two examples for which eye detector fails

4. CONCLUSION

In this paper we proposed a method for detection of eyes in 2D greyscale of human face. We described the face image and the model eyes using the topographic labels. The regions provided from this labelling are then describe using statistical moments which are invariant to similarity transformation. Using the topographic labelling made the proposed method robust to illumination change. The experimental results also show that the proposed method is invariant to head rotation. This property is result of invariance of employed descriptors to similarity transform.

REFERENCES

[1] H. Tan, Y. j. Zang, and R.Li, "Robust Eye Extraction Using Deformable Template and Feature Tracking Ability,"in *Proc.* ICICS-PCM, 15-18 December 2003, vol. 3, pp. 1747 - 1751.

[2] H. Gu, G. Su, and C. Du. "Feature Points Extraction from Faces", in *Proc Image and Vision Computing*. New Zealand, 26-28 November, 2003,pp. 154-158.

[3] L.Jin, X.Yaun, S.Satoh, J.Li, and L.Xia. "A hybrid classifier for precise and robust eye detection," in *Proc.* ICPR'06, 2006, vol. 4, pp. 731-735.

[4] A. Fathi, and M. T. Manzuri, "Eye Detection and Tracking in Video Streams", in *Proc* .ISCIT 2004, 26-29 Octobr 2004, pp. 1258-1261.

[5] G. Marcone, G. Martinelli, and L. Lancetti, "Eye Tracking in Image Sequences by Competitive Neural Networks," *Springer Journal: Neural Processing Letters,* vol. 7, No. 3, pp. 133-138, Jun 1998.

[6] J. Wang, and L. Yin." Detecting and Tracking Eyes Through Dynamic Terrain Feature Matching," in *Proc.* CVPR'05, 2005, Vol. 3, 20-26 June, pp. 78 - 78.

[7] R. M. Haralick. L. T. Watson, and T. J. Laffey, "The topographic primal sketch," *Int. J . Robotics Res.* vol. 2, pp. 50-72, 1983.

[8] J. Wang, and L. Yin, "Static topographic modeling for facial expression recognition and analysis," *Computer Vision and Image Understanding Journal,* vol. 108, pp. 19-34, October 2007.

[9] L. Wang, and T. Pavlidis."Direct Gray-Scale Extraction of Features for Character Recognition",*IEEE TRANSACTIONS On Pattern Analysis and Machine Inter*
face. vol. 15, no. 10, pp.1053-1067, October 1993.

[10] P. Meer and I. Weiss, "Smoothed differentiation filters for images," in *proc*., 10th International Conference on Pattern Recognition, Maryland, USA, 16-22 Jun, 1990, vol. 2, pp. 121-126.

[11] M. Hu., "Visual Pattern Recognition by Moment Invariants,". *IRE Transactions on Information Theory*, pp. 179-187, 1962.