



ELSEVIER

Pattern Recognition Letters 21 (2000) 463–471

Pattern Recognition  
Letters

www.elsevier.nl/locate/patrec

# A new face detection method based on shape information

Jianguo Wang<sup>\*</sup>, Tieniu Tan

*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, People's Republic of China*

Received 5 October 1998; received in revised form 17 November 1999

## Abstract

Automatic detection of human faces is one of the most difficult problems in pattern recognition. In many practical applications (e.g. personal identification using photo-IDs), face images often have a simple background. In this paper, we propose a new method based on shape information. Our system can be very efficient for images with simple background. Input image is first enhanced by means of histogram equalization, followed by edge detection based on a multiple-scale filter. The extracted edges are then linked using a method based on an energy function. The face contour is finally extracted using the direction information of the linked edges. Experimental results are included to verify the effectiveness of this method. © 2000 Elsevier Science B.V. All rights reserved.

*Keywords:* Face detection; Image segmentation; Template matching

## 1. Introduction

Face detection is the important first step of face recognition. It is basically an image segmentation problem as the image is to be segmented into two parts: one containing faces and the other representing non-face regions. The performance of face detection influences the performance of the entire recognition system.

A number of methods have been proposed, and some have been used in practical applications (Miller, 1994). These methods can be broadly classified into two groups: the first group is based

on separate facial features, which means the final decision comes from the integration of several detection results. Among these, Yang and Huang have described a system that uses a hierarchical knowledge-based method; Li and co-workers (Chow and Li, 1993; Li and Roeder, 1995) use several deformable templates to locate the face contour; facial feature detection and verification based on geometrical relations have been studied in (Jeng et al., 1996). The second group considers a face as a single pattern and features are extracted from the entire face region. Existing methods have employed techniques ranging from neural networks (Rowley et al., 1998a,b; Juell and Marsh, 1996; Lin and Kung, 1997; Ranganathan and Arun, 1997), color information (Saber and Tekalp, 1996; Sobottka and Pitas, 1996) to spatial gray-level dependence matrix (SGLD) (Dai and Nakano, 1996) etc. In most cases, multiple techniques are used together to achieve better performance.

<sup>\*</sup>Corresponding author. Present Address: Department of Computer Science, Michigan State University, 3115 Engineering Building, East Lansing, MI 48824, USA. Tel.: 517-353-9574; fax: 517-432-1061.

*E-mail addresses:* wangjia3@msu.edu (J. Wang), tnt@nlpr.ia.ac.cn (T. Tan).

We realize that face images with simple background are seen in many practical applications, such as photo-IDs for personal identification and security work. So far several different methods have been proposed toward such specific images. Among these, the method of template matching (Yang and Huang, 1994; Chow and Li, 1993) is most prevalent. In this method, the face contour is extracted using six templates—Two eye templates and one mouth template are used to verify a face and locate its main features; then two cheek templates and one chin template to extract the face contour. The performance is favorable except that it cannot detect faces with shadow, rotation and bad lighting conditions.

In most cases, the overall shape of the face or the whole head is very similar to an ellipse. Several methods have been proposed using this shape information (Saber and Tekalp, 1996; Tang et al., 1998; Sobottka and Pitas, 1996). However, face or head contour cannot easily be modeled using a single ellipse. In this paper, we attempt to extend previous shape-based face detection methods by applying a special template containing directional information of edges. Extensive experiments show that the new system is very efficient when processing images with a simple background regardless of variations on size, head pose (moderate head rotation) and lighting condition.

Details of this system are described in the remainder of this paper. Section 2 covers the basic principles of the system. In Section 3, we give the experimental results and the detailed comparison with other systems. Section 4 concludes the paper with suggestions for possible extension and improvements.

## 2. Description of the system

The basic flowchart of our system is given in Fig. 1. For the first three steps, we only give the basic principles for they can be easily found in many image processing textbooks. The last two steps are the kernel of our method and we discuss them in greater details.

### 2.1. Image enhancement

Input images may be of very poor contrast because of the limitation of lighting conditions. In our system histogram equalization is used to improve the contrast of the original image.

Fig. 2 gives an example to illustrate the effectiveness of image enhancement. Fig. 2(a) is the original image of bad contrast; Fig. 2(c) is the output; Fig. 2(b) and Fig. 2(d) are the edge images,

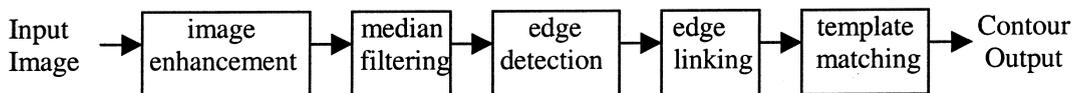


Fig. 1. Basic flowchart of the algorithm.

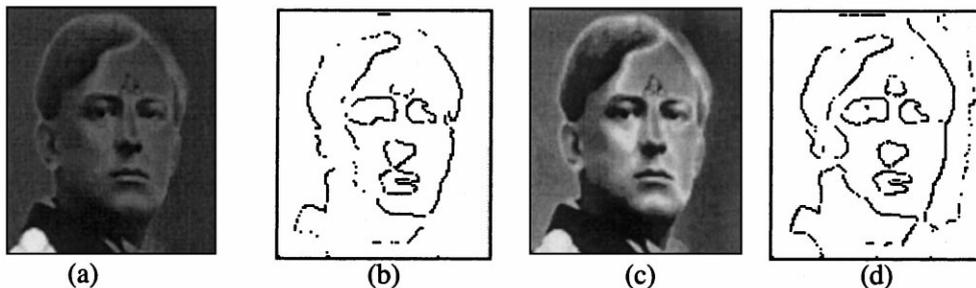


Fig. 2. Effectiveness of histogram equalization.

respectively, extracted from Fig. 2(a) and Fig. 2(c) using the method illustrated in Section 2.3.

Comparing Fig. 2(b) with Fig. 2(d), it is obvious that the number of points on the face contour has been increased, which means that facial information has been strengthened. Noise has also been enhanced but in the subsequent steps it can be erased.

### 2.2. Median filtering

Various sources of noise exist in the input image. The fine details of the image represent high frequencies which mix up with those of noise. When low-pass filtering is used, some details in the image may be erased as well. In this paper, we use median filtering to remove noise.

Fig. 3 illustrates the effectiveness of median filtering. Fig. 3(a) is the output of Fig. 2(c) after median filtering; Fig. 3(b) is the edge image extracted from Fig. 3(a).

Comparing Fig. 3(b) with Fig. 2(d) we note that noise has been reduced with little change of the face contour.

### 2.3. Edge detection

There are many edge detectors to choose. Considering computational cost and performance, we choose the zero-crossing detector: Difference of Exponential (DOE) (Ma and Zhang, 1998)

$$DOE = K_c(e^{-|x|/t_1} - e^{-|x|/t_2}). \quad (1)$$

This operator can be implemented recursively, so its computational cost is very low. In (Ma and

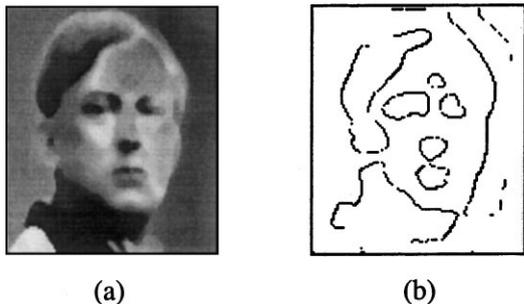


Fig. 3. Effectiveness of median filtering.

Li, 1995), it has been proved that with the scale ratio (denoted as  $t_1/t_2$ ) of 0.3 this edge detector has optimal performance. And when the scale is bigger only the edge of relatively larger region can be detected. In this paper we choose  $t_1 = 1.5, t_2 = 5.0$ .

### 2.4. Edge linking

In (Ke et al., 1996), the authors propose the following energy function to represent the global saliency of a solitary line contour  $l_{(s)} = (x_{(s)}, y_{(s)})$ , where  $(x, y)$  is the Euclidean coordinate. The energy function  $H$  is defined as

$$H = 2h_0 + \frac{1}{2} \int k_s (\nabla l_s)^2 ds + \frac{1}{2} \int b_s (\nabla^2 l_s)^2 ds, \quad (2)$$

where  $s$  is the arc length parameter ranging on  $[0, L]$  with  $L$  being the length of the line contour,  $k_s$  and  $b_s$  are elastic coefficients. We denote  $\nabla l_s = (\partial x / \partial s, \partial y / \partial s)$  and  $\nabla^2 l_s = (\partial^2 x / \partial^2 s, \partial^2 y / \partial^2 s)$ . In Eq. (2), the first term represents the energy for two end points of  $l_s$ ; the second and the third term are due to stretching and bending modes, respectively. In the experiments,  $h_0$  is a constant to represent the base of the energy function and the other two terms reflecting the length and the curvature property of the line. The longer the line and the smaller the curvature, the smaller is  $H$  and the more salient the line contour. More details concerning Eq. (2) may be found in (Ke et al., 1996). This method has been effectively used in roads and blood vessels extraction. The linking process (grouping) consists of the following four steps:

1. Forming the edge chains from the input edge image;
2. Erasing the chains whose length is shorter than  $a$  to decrease noise. In general  $a$  ranges from 1 to 5; here we choose  $a = 3$ ;
3. Decomposing each line contour into a set of segments with constant curvature: straight lines and arcs; storing them into a new chain;
4. For every segment, the other one is searched in a defined region to make their linkage have optimal global saliency.

After linking the function  $H$  can achieve optimal value. Fig. 4 is the output of Fig. 3(b) after



Fig. 4. Edge linking.

linking. The edges have been linked well so the information of the face contour is improved. Also noises have been greatly reduced.

### 2.5. Template matching

In many papers (Chow and Li, 1993; Li and Roeder, 1995; Jeng et al., 1996; Kwon and Labo, 1994), various templates have been proposed. Here, we present an original template based on edge direction. It has been noticed that the contour of a human head can be approximated by an ellipse. This accords with our visual perception and has been verified by numerous experiments (Tang et al., 1998). The existing methods have not sufficiently used the global information of face images in which edge direction is a crucial part, so we present a deformable template based on the edge information to match the face contour.

The face contour is of course not a perfect ellipse. To achieve good performance, the template must tolerate some deviations. In this paper, we

use an elliptical ring as the template as illustrated in Fig. 5.

Fig. 5(a) is a normal upright face; (b) is the binary image after edge linking. In (b), we can note that the external contour cannot be represented by a single ellipse no matter how the parameters are adjusted. However, if we use an elliptical ring representing the contour as shown in Fig. 5(c), almost all the edge points on the contour can be included. The other important advantage of this template is that we can choose a relatively big step in matching so as to reduce the computational cost.

An ellipse is described by the following equation:

$$\frac{(x - X_0)^2}{a^2} + \frac{(y - Y_0)^2}{b^2} = 1, \quad (3)$$

where  $(X_0, Y_0)$  is the center of the ellipse,  $a$  and  $b$  are the long and the short axis, respectively. If we let  $b = C_1 a$ , with  $C_1$  a constant, then  $(X_0, Y_0, a, C_1)$  are the four free parameters that describe the upright ellipse as displayed in Fig. 6. In this figure,  $A$  is a random point on the ellipse;  $AB$  is the tangent intersecting the positive  $X$ -axis at  $B$ . We denote the tangent angle as  $\theta$  computed by:

$$\theta = \arctan \left( -C_1^2 \times \frac{x - X_0}{y - Y_0} \right). \quad (4)$$

In this paper, we use “inner product” of two vectors to evaluate the difference of the direction of the edge and the template. This is better than the absolute difference of two angles used in

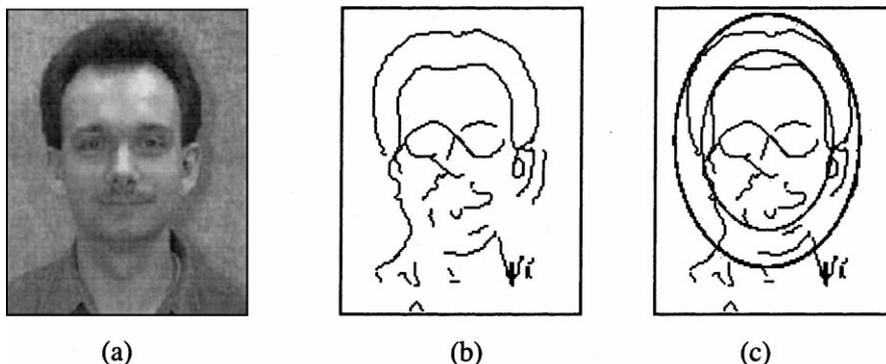


Fig. 5. Deformable template.

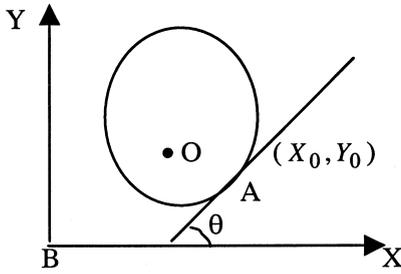


Fig. 6. An ellipse in 2-D space.

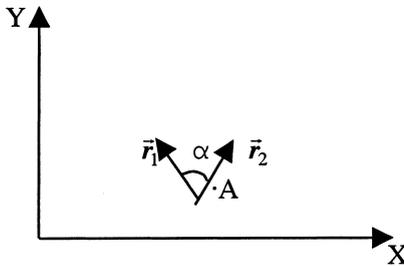


Fig. 7. Vector space.

(Wang and Tan, 1999). In the vector space as displayed in Fig. 7, A is the point shown in Fig. 6.  $\vec{r}_1$  is the unit vector of the tangent of edges at A, while  $\vec{r}_2$  is the unit vector of the tangent of the template at A.  $\alpha$  denotes the angle of the two vectors. Then we have

$$\vec{r}_1 \bullet \vec{r}_2 = \cos(\alpha). \quad (5)$$

The value of the inner product can effectively reflect the difference of the two vectors. In our experiments, we first calculate the gradient direction of each edge point with the Sobel operator. Then the expression of  $\vec{r}_1$  is derived. The expression of  $\vec{r}_2$  is easily derived from the value of  $\theta$  as  $\vec{r}_2 = \cos(\theta)\vec{i} + \sin(\theta)\vec{j}$ , where  $\vec{i}$  and  $\vec{j}$  are the unit vectors along the X and Y axis, respectively.

For the evaluation of the goodness of match between the extracted face contour and the template, we denote the sum of the inner product and the number of edge points in a template as  $G$  and  $N$ , respectively, and define the evaluation score  $R$  as

$$R = \frac{G}{N}. \quad (6)$$

The denominator  $N$  is to normalize large elliptical templates because the larger the template the higher the value of  $G$ . This evaluation function has been verified by our training sets. In the process of confirming candidate faces, two thresholds are used to decrease computational cost:

1.  $N_t$ ; only when the number of edge points ( $N$  in Eq. 6) on the template is above  $N_t$ , we continue the confirmation step;
2.  $R_t$ ; in the confirmation step, when the value of  $R$  is above  $R_t$  a candidate face is confirmed.

In our experiments, these thresholds are chosen by training on a set of representative images and then kept unchanged. The matching algorithm can now be summarized as follows:

1. For each set of temple parameters  $(X_0, Y_0, a, C_1)$
2. Count the edge points lying on the template (denoted as  $N$ )
3. If  $N < N_t$ , then go to 1
4. Else calculate the value of  $R$
5. If  $R < R_t$ , then go to 1
6. Else store the value  $R$  with the template parameters
7. The parameter set with the largest evaluation score is chosen as the final result.

### 3. Experimental results and comparison

In experiments, a set of 20 images is used as the training set. All thresholds and parameters used in our system are obtained from the training set.

For clarity, we define the following terms:

*Correct detection*: refers to the detection results which can be used in applications such as recognition, tracking etc.

*False detection*: for detection results not classified as correct detection.

First our algorithm is tested on MIT face database. Each of 16 people is digitized 27 times in the office, varying the head orientation, the lighting, and the scale (camera zoom). The resolution is  $128 \times 120$  with moderate complicated background. Table 1 gives detailed detection result and Fig. 8 shows a group of examples.

Background of MIT images is single. Another set containing 90 images is set up to give a more convincing test. Some of the test images came from

Table 1  
Detection results on MIT database

	Correct detection		False detection	
	Inexact <sup>a</sup>	Exact <sup>a</sup>		
Our system	50/432	11.57%	84.96%	15/432 3.47%

<sup>a</sup> Inexact means that the ring has passed through some part of the contour and face is mainly included, but non-face region also included or some facial features missed, so it is hard used for identification; Exact is the other kind in “Correct detection”. Inexact results are mainly caused by tilted faces (22.5°). The database has 144 upright and 288 tilted faces, which contributes the high inexact rate.

the CMU face database and World Wide Web; others were obtained from a digital camera (office background, resolution 1280 × 960) or scanned

from outdoor photographs (resolution 600 dpi). Here, all of our test images contain only one face (multi-face detection is discussed in Section 4). The size, lighting condition and head pose are different.

The test set is decomposed into two subsets: one containing 50 images with simple background, the other containing 40 images with complex background.

In the first subset, the system can give the correct locations for all 50 images. But 8 of them cannot give exact detection because of beards, large tilting angles etc. The rate of correct detection is 100% in which inexact detection rate is 16%. The false rate is 0. The performance is very good. Some examples with simple backgrounds are given in Fig. 9.

In the second subset, the system can accurately locate 87.5% of the faces (35 from 40) in which 5



Fig. 8. Detection results of MIT images.



Fig. 9. Face detection of images with simple backgrounds.

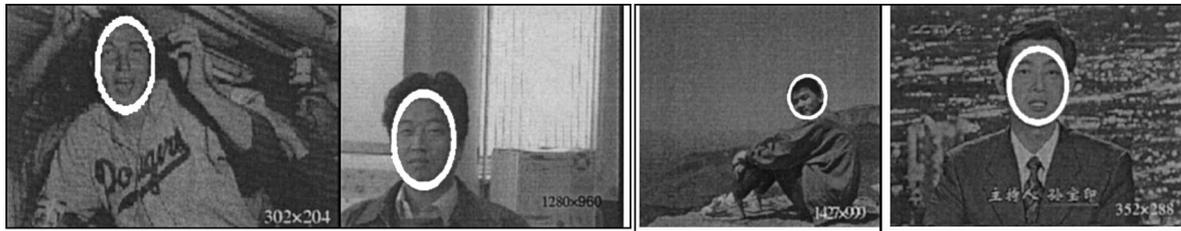


Fig. 10. Face detection of images with complex backgrounds.

elliptical rings give inexact locations (the rate is 12.5%); other 5 results have no relations with the faces (the false rate is 12.5%). The disturbance of so many irrelevant edges of the background causes the high rate of false detection. Four examples in Fig. 10 are chosen from the CMU database, showing images with office background, outdoor scene and TV video, respectively.

Govindaraju (1996) also described a system based on the shape information to locate faces. After edge filtering and linking, a set of three special strings ( $L, H, R$ ) is used to model a face.  $L, H, R$ , respectively, correspond to the left side, the hair-line and the right side of the front view of a face. The attributes mainly contain the orientation, length, continuity and their mutual relationship. For each candidate face a cost reflecting the distance with an ideal face is assigned. Candidate locations whose cost is below a pre-defined threshold are selected as the final faces. Their system can also be used for multi-face detection where the number of people is obtained from the photo caption. But in the experiments, the three strings cannot uniquely define a face. Although some approaches have been described, false alarms still exist more than the user can accept. The system also requires that each string be uninterrupted. Although edge has been linked, this requirement is difficult to meet. When processing faces with pose changes, false alarms and inexact locations increase greatly. Our system searches one optimal location in the image (for multi-face detection, refer to Section 4). The template aims to grasp the global information of the object and requires no detailed information about separate face segments. Therefore for images with variable head poses and bad contrast,

Table 2

Comparison between Govindaraju, 1996 and our system

Performance system	Correct detection rate <sup>a</sup>	For faces with shadow, tilting or bad contrast <sup>b</sup>
Govindaraju, 1996	83%	Good
Our system	87.5%	Better

<sup>a</sup> The rate of Govindaraju, 1996 is based on one of his test set where each image contains one face. Although the two systems are tested on different databases the face images used are comparable.

<sup>b</sup> In (Govindaraju, 1996), inexact detection rate is higher and more faces are missed than in our system.

the algorithm described in this paper is expected to outperform that of Govindaraju, 1996. The comparison in Table 2 shows that in addition to higher detection rate, our system is also more robust to noise.

Although our system is more robust than Govindaraju, 1996, the template described here cannot uniquely model a face either, especially when the images have objects whose shape is similar to ellipses (though these cases are not typical in real images). The key idea to remove the false locations has been discussed in Section 4.

In (Saber and Tekalp, 1996), a system integrating color, shape and symmetry information has been proposed. The ellipse is used only as an auxiliary step to remove the skin regions other than faces. Compared with our system, their face model is very simple but enough to delete the wrong regions in the candidate space. Better performance may be achieved after color information is integrated into our system. This is discussed as our future work in Section 4.

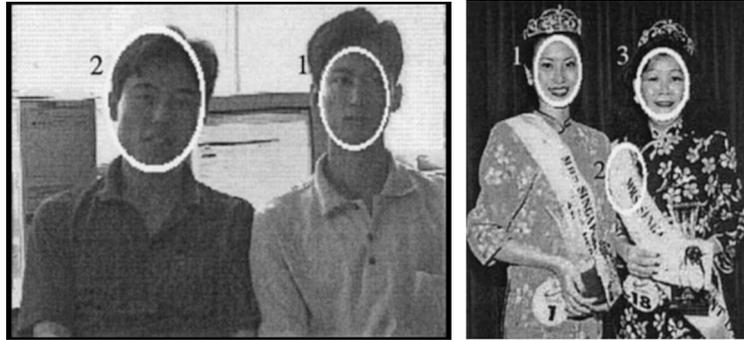


Fig. 11. Multi-face detection.

#### 4. Extensions and conclusions

This paper has demonstrated the effectiveness of a new face detection algorithm in the images with simple or complex background. The algorithm is able to correctly detect all faces in the images with simple backgrounds. Compared with other similar algorithms, the new algorithm appears to be more robust to noise and shape variations.

We intend to extend the algorithm for multi-face detection. First, the number of faces should be known before the detection. This is similar to Govindaraju (1996) who obtains the number information from picture captions. Second, we assume that faces do not overlap each other in images. Then after matching, we select a fixed-number of locations with high evaluation score as face regions. If two are overlapped, the detection with lower confidence is substituted by another one with highest score in the waiting list.

Fig. 11 shows some preliminary results. The number near the face indicates the order of confidence after matching. The left image (in which the left face is a little tilted) is obtained from a camera and the result is very good. The right one with various colorful clothes is from the Internet. As can be seen, the left face can be detected first, then a false alarm and then the right face. That is, when we set the face number as three, two faces can be detected with a false alarm (near the center of the image). We suspect that one reason for this is that our template does not include enough information to distinguish faces in very complex backgrounds.

This is also why the false rate is high in our test set with complex backgrounds.

For future work, we have several directions. The computational cost can be reduced further. Other information can be integrated to identify candidate regions (such as color or symmetry). Another limitation of the system is that it cannot accurately locate faces with large rotation angles. This is due to the fact that for speed considerations, there is no angle parameter in the ellipse equation. However, if candidate regions can be identified, the angle could be included. Finally in order to improve the rate of correction detection on complex backgrounds, a verification step should be added to remove false locations.

#### References

- Chow, G., Li, X., 1993. Towards a system for automatic facial feature detection. *Pattern Recognition* 26 (12), 1739–1755.
- Dai, Y., Nakano, Y., 1996. Face-texture model based on SGLD and its application in face detection in a color scene. *Pattern Recognition* 29 (6), 1007–1017.
- Govindaraju, V., 1996. Locating human faces in photographs. *Internat. J. Comput. Vision* 19 (2), 129–146.
- Jeng, S.H., Liao, H.M., Liu, Y.T., Chern, M.Y., 1996. An efficient approach for facial feature detection using geometrical face model. In: *Proceedings of the ICPR 1996*, pp. 426–430.
- Juell, P., Marsh, R., 1996. A hierarchical neural network for human face detection. *Pattern Recognition* 29 (5), 781–787.
- Ke, Q., Xiao, J., Yang, Z., 1996. Energy-based Method for Road Extraction from Satellite Images. In: *Proceedings of the IAPR Workshop on Machine Vision Application*, pp. 337–340.

- Kwon, Y.H., Lobo, N.V., 1994. Face Detection Using Templates. In: Proceedings of the ICPR, pp. 764–767, (Ch. 200).
- Li, X., Roeder, N., 1995. Face contour extraction from front-view images. *Pattern Recognition* 28 (8), 1167–1179.
- Lin, S.H., Kung, S.Y., 1997. Face recognition/detection by probabilistic decision-based neural network. *IEEE Trans. Neural Network*, Vol. 8, No. 1.
- Ma, S., Li, B., 1995. Derivative Computation by Multiscale Filters. In: Proceedings of Europe-China Workshop on Geometric Modeling and Invariant, Xi'an, China.
- Ma, S., Zhang, Z., 1998. *Computer Vision*, Chinese Scientific Press.
- Miller, B., 1994. Vital signs of identity. *IEEE Spectrum* 2, 22–30.
- Ranganath, S., Arun, K., 1997. Face recognition using transform features and neural network. *Pattern Recognition* 30 (10), 1615–1622.
- Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural Network-based Face Detection. *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 20, No. 1.
- Rowley, H.A., Baluja, S., Kanade, T., 1998. Rotation Invariant Neural Network-based Face Detection. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition.
- Saber, E., Tekalp, A.M., 1996. Face detection and facial feature extraction using color, shape and symmetry-based cost functions. In: Proceedings of the ICPR, Vol. 3, pp. 654–658.
- Sobottka, K., Pitas, I., 1996. Extraction of facial regions and features using color and shape information. In: Proceedings of the ICPR, Vol. 3, pp. 421–425.
- Tang, C., Hung, Y., Chen, Z., 1998. Automatic detection and tracking of human heads using an active stereo vision system. In: Proceedings of the ACCV, Vol. 1, pp. 632–638.
- Wang, J., Tan, T., 1999. Face Detection Using Shape Information – an Efficient Method for Images with Simple Background. In: Proceedings of the The Second International Conference on Multimodal Interface, Hong Kong.
- Yang, G.Z., Huang, T.S., 1994. Human face detection in a complex background. *Pattern Recognition* 27 (1), 53–63.