Gazing-detection of human eyes based on SVM*

LI Su-mei**, **ZHANG Yan-xin**, **CHANG Sheng-jiang**, and **SHEN Jin-yuan**
Institute of Modern Optics, and Key Laboratory of Opto-electronics Information Technical Science, EMC, Nankai University, Tianjin, 300071, China

(Received 20 April 2005)

A method for gazing-detection of human eyes using Support Vector Machine (SVM) based on statistic learning theory (SLT) is proposed. According to the criteria of structural risk minimization of SVM, the errors between sample-data and model-data are minimized and the upper bound of predicting error of the model is also reduced. As a result, the generalization ability of the model is much improved. The simulation results show that, when limited training samples are used, the correct recognition rate of the tested samples can be as high as 100%, which is much better than some previous results obtained by other methods. The higher processing speed enables the system to distinguish gazing or not-gazing in real-time.

CLC number: TP391.41 **Document code**: A **Article ID**: 1673-1905(2005)01-0065-04

Gazing-detection plays an important role in natural interaction among individuals which usually reflects the intention of both sides. The detection ability of vision for gazing and facial expression perception is very unique, especially for its almost infinite generalization. Therefore, research to understand its mechanism has been a frontier in vision science. Moreover, with the development of robot and human-machine-interaction (HMI), the contact-free HMI by using speech and gesture have received increasing interest of research in recent years. Especially, gazing-detection is becoming more and more important not only in theoretical study but also in a variety of applications. There is a great deal of researches on the related subject undertaken both at home and abroad. P. W. Hallinan[1] proposed to use a series of parameters to determine human eyes on the basis of its geometry characters; A. Yuille[2] Jyh-yuan Deng[3] and X. Xie[4] proposed to use changed templates and to construct cost functions for determining the orientation of eyes and extracting parameters. A. Nikolaidis[5] used the symmetry of facial characters to gain the angle between facial plane and image plane. In all these approaches, however, the user had to stay in an almost fixed position and were not allowed to turn their heads, and a special lighting was needed as well. Gee and Cipolla[6] developed a system to track the rotation and positions of a head by finding correspondences between the facial feature points and the points in a model of the head, using a weak perspective projection. However, the system had to be initialized manually because it could not locate the face and the facial feature points automatically. K. Talmi and J. Liu[7] presented a video-based contact-free measurement system which allowed combined tracking of subject's eye position and gaze direction in quasi-real time. But three cameras were needed in their system, and the eye image must be of high resolution. Wang Yong et al[8,9] used logic analysis and a simple two-layer neural network to determine gazing or not gazing based on the geometry parameters for eye center and eyeball center, and got some useable results. Up to now, there have been many technical factors, such as requirements to the lighting and recording resolution, distance and the numbers of persons being detected, the accuracy of gazing-detection and the speed of gazing tracking, which still restrict the applications of the gazing-detection. Therefore, further research on the techniques have been urgently needed to ensure practical applications.

Apart from the research of using a lot of assistant instruments[10] to measure accurately the gazing direction, which may be called "gaze measurement", we call the other research "gazing- detection" which focused on whether or not a person is gazing at some specific spot.

Support vector machine (SVM), pioneered by Vapnik, 1992 and rooted in the statistical learning theory, is a machine learning algorithm. More precisely, the SVM is an approximate implementation of the method of structural risk minimization. The errors between sample-data and model-data are minimized and the upper bound for predicting error of the model also decreases simultaneously, therefore the ability of generalization of the model is much improved. However, the traditional learning method (e. g. BP neural networks) is an approximate implementation of the method of empirical risk minimization with higher learning precision results in worse ability of generalization, which is so called "overfitting". Obviously, the feature of SVM is very important for the research of gazing-detection. In this paper, the SVM method is used instead of BP networks. The simulation results show that the correct recognition rate of the testing samples can be as high as 100%, which is

---

* Supported by National Natural Science Foundation of China (No. 60478036)

** E-mail: hbu lisumei@sina. com

much better than some previous results obtained by other methods. The higher processing speed enables the system distinguishes the gazing direction in real-time, and its property of generalization has approached to that of gazing-detection of human vision.

The main idea of a support vector machine is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized. Fig. 1 illustrates the geometric construction of an optimal hyperplane for a two-dimensional input space. The separation between the hyperplane H and the closest data point is called the margin of separation, denoted by ρ.

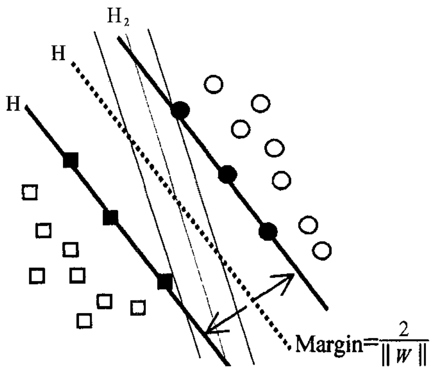


Fig. 1 Illustration of the idea of an optimal hyperplane for linearly separable patterns

For the linearly separable samples data: $(x_i, d_i), i=1,2,\cdots,n, x\in R^d, d\in\{+1,-1\}$, the equation of a decision surface is $w\cdot x+b=0$, when the constraint $d_i[(w\cdot x)+b]-1\geqslant 0, i=1,2,\cdots,n$ is satisfied, the margin of separation is equal to $2/\|w\|$. Therefore, to maximize the margin of separation is to minimize the $\|w\|^2$, that is, to optimize the following equation:

$$\begin{cases}\min Q(w) = \frac{1}{2}w^T w \\ d_i[(w\cdot x)+b]-1\geqslant 0, i=1,2,\cdots,n\end{cases} \tag{1}$$

with the Lagrange multipliers, we transform it to dual problem which has the same optimal value as it.

$$\begin{cases}\max Q(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j d_i d_j (x_i\cdot x_j) \\ \sum_{i=1}^{n} d_i\alpha_i = 0, \alpha_i\geqslant 0, i=1,2,\cdots,n\end{cases} \tag{2}$$

The auxiliary nonnegative variables α_i are called Lagrange multipliers, and this dual problem's solution is unique. Corresponding α_i of the solution for the dual problem is not zero (usually they are few), corresponding samples of α_i are support vectors: SVS. Its optimal separable function is:

$$f(x) = \mathrm{sgn}\{(w\cdot x)+b\} = \mathrm{sgn}\left[\sum_{i=1}^{n}\alpha_i^* d_i(x_i\cdot x)+b^*\right] \tag{3}$$

α_i^* is the Lagrange multiplier corresponding SVS, to compute the bias b^*, we may use the w_0 thus obtained or take advantage of any support vector to get it.

For the nonlinearly separable sample data, some points of the data (x_i, d_i) may fall on the wrong side of the decision surface. To decrease the risk of error, a new set of nonnegative scalar variables $\zeta_i\geqslant 0$ is introduced into (1) as shown below:

$$d_i[(w\cdot x_i)+b]-1+\zeta_i\geqslant 0, i=1,2,\cdots,n$$

where ζ_i are called slack variables. So the objective function is as follows:

$$Q(w,\zeta) = \frac{1}{2}w^T w + C\left|\sum_{i=1}^{n}\zeta_i\right| \tag{4}$$

We minimize the objective function (4) with respect to the weight vector. The parameter C controls the tradeoff between complexity of the machine and the number of nonlinearly separable points. The value of parameter C has to be selected by the user. The nonlinearly separable case differs from the nonlinearly separable case in that the constraint $\alpha_i\geqslant 0$ is replaced with the more stringent constraint $0\leqslant\alpha_i\leqslant C, i=1,2,\cdots,n$. Except for this modification, the constrained optimization for the nonlinearly separable case and computations of the optimum values of the weight vector W and bias b proceed in the same way as in the linearly separable case.

In fact, many problems of application are nonlinearly separable. We may map the input vector into a higher-dimensional feature space, then, the optimal decision surface is obtained in higher-dimensional. According to functional analysis, if the kernel function satisfies the Mercer's theorem, it may become the inner-product in some space. The transform $\Phi: x\rightarrow z$ map the input vector space into a feature space. $K(x_i, x_j)$ is the inner-product in feature space, that is, Kernel function:

$$\langle\Phi(x_i),\Phi(x_j)\rangle = K(x_i, x_j) \tag{5}$$

So its dual problem is shown as follows:

$$\begin{cases}\max Q(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j d_i d_j K(x_i, x_j) \\ \sum_{i=1}^{n}\alpha_i d_i = 0, 0\leqslant\alpha_i\leqslant C, i=1,2,\cdots,n\end{cases} \tag{6}$$

and its optimal separable function is:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^* d_i K(x_i, x) + b^* \right) \quad (7)$$

Fig. 2 displays the architecture of a support vector machine. The support vector machine model is similar to RBF neural networks model, every node in hidden layer is corresponding with a support vector. The difference between SVM and RBF is as follows: In SVM the number of radial-basis functions and their centers are determined automatically by the number of support vectors and their values, respectively. However, in RBF they are determined by the user's experience or learning.

The features related to gazing detection are hidden in the segmented image in which there are noises introduced by the camera itself, lighting, and so on. Therefore, it is necessary to preprocess the image data to extract or emphasize the related features. In our experiment, we firstly segment the eye window from the segmented facial image. By analyzing and comparing their histogram, we choose their red components (x) as the features for gazing detection. Then we transform the red components nonlinearly into $y = \exp(-\alpha x)$. Fig. 3 displays the process of preprocessing. It has been shown that the preprocessing can avoid the errors caused by extracting the facial geometry parameter as did in reference[8,9] and reduce the time of process.

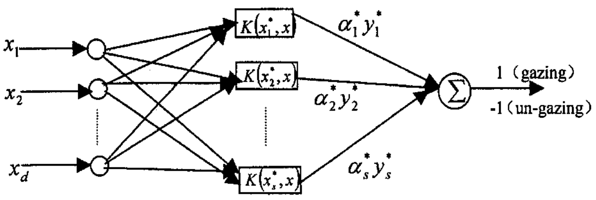
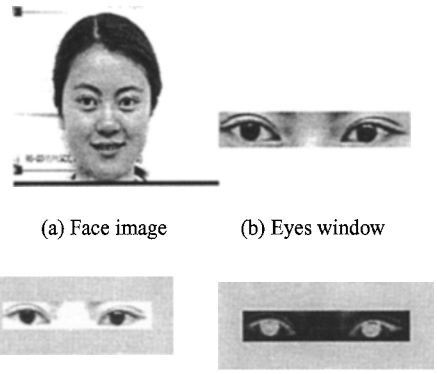


Fig. 2 The network model of SVM

In our experiment, the gazing and un-gazing images of 10 persons who are facing to the camera are taken and digitalized into a computer. The number of gazing samples is 109, and so is of the un-gazing ones. Among them 128 samples were used for training by using SVM, and 90 remnants for test. The resulted rate of correct recognition for the test samples can be as high as 100%. Tab. 1 displays correct recognition rate for different models. From it the generalization and practicability of different models can be compared. PAN in Tab. 1 denotes the "parameter analysis-NN" model for reference [8, 9]; GPN in Tab. 1 denotes "geometry parameter-NN" model; P-SVM is the "preprocessing-SVM" model mentioned in this paper. Dispite the correct recognition rate of 95% for GPN, this is the result for all training samples(200), the number of its testing samples is 0. So the generalization is almost 0. In the PAN model, its network structure for 4-2-1 is simple, and delete many samples artificially, but its correct recognition rate is far lower than P-SVM, and it is not in accord with vision mechanism of human eyes. So, for gazing detection, P-SVM model enables the system distinguishes the gazing direction in real-time, as well as to better approach to the characteristics of gazing detection of human vision.



(a) Face image (b) Eyes window

(c) Red component of eyes window (d) Eyes window of nonlinearly preprocessing

Fig. 3 Illustration of data preprocessing

Tab. 1 Comparision of correct Recognition Rate for Different Model

Different models	Total number of samples	Number of training samples	Number of testing samples	Rate of correct recognition	Training time
PAN[6,7]	253	20	27	85%	No applicable
GPN	200	200	0	95%	10-20 h
P-SVM	218	128	90	100%	3-5 min

This paper describes a real-time method for gazing-detection based on the method of SVM, in which the datum pre-processing to emphasize the features is used and no special illumination and headgear are needed. According to the criteria of the minimization for the structural risk of SVM, the errors between sample-data and model-data are minimized and the upper bound for predicting error of the model is also decreased simultaneously, therefore the ability of generalization of the model is much improved. The simulation results show that higher processing speed, better correct recognition rate, and improved generalization can be obtained by using of this method. The higher processing speed enables the system to distinguish between gazing or not-gazing in real-time, and the improved property of generalization has approached to that of gazing detection of human vision.

References

[1] P. W. Hallinan, *SPIE* **1570**, (1991), 214.
[2] A. Yuille, D. Cohen, and P. Hallinan. *Proc. CVPR*, 1989, 104.

[3] Jyh-yuan Deng and Feipei Lai. Pattern Recognition, **30** (1997), 403.

[4] X. Xie, R. Sudhakar, and H. Zhuang. Pattern Recognition, **27** (1994), 791.

[5] Athanasios Nikolaidis and Ioannis Pitas. Pattern Recognition, **33**(2000), 1783.

[6] A. Gee and R. Cipolla. *Proc. Mechatronics and Machine Vision in Practise*, Toowoomba, Australia, (1994), 112.

[7] K. Talmi and J. Liu. *Signal Processing: Image Communication*, **14**(1999), 799.

[8] WONG Yong, YUAN Jing-he, CHANG Sheng-jiang, and ZHANG Yan-xin. *Journal of Optoelectronics • laser*, **12**(2001), 1284. (in Chinese)

[9] WANG Yong, YUAN Jing-he, CHANG Sheng-jiang, and ZHANG Yan-xin. *Optical Engineering*, **41**(2002), 1840.

[10] D. A. Simon, M. Hebert, and Kanade. *International Conference of Robotics and Automation Proceeding*, May'94, San Diego, USA.