Journal of Pattern Recognition Research 1 (2006) 55-62

Eye Detection in Facial Images with Unconstrained Background

Qiong Wang *nustdaisy@gmail.com*
Jingyu Yang *yangjy@mail.njust.edu.cn*
School of Computer Science & Technology, Nanjing University of Science & Technology
200 Xiaolinwei Street, Nanjing, Jiangsu Province, 210094, China

Received Jun. 2, 2006. Revised Sep. 12, 2006. Published Sep. 25, 2006.

Abstract

This paper presents an efficient eye detection approach for still, grey-level images with unconstrained background. The structure of the eye region is used as a robust cue to find eye pair candidates in the entire image. Eye pairs are located by a support vector machine-based eye verifier. The eye variance filter is then used to detect two eyes in the eye region which has been extracted in the eye pair location step. The proposed method is robust against clustered background, moderate rotations, glasses wearing, and partial face occlusions. The method is evaluated using the BioID face database. The experimental results demonstrate the effectiveness of the presented method.
Keywords: Eye detection, Binary template matching, SVM, Variance filter.

1. Introduction

Eye detection is a crucial aspect in many useful applications ranging from face recognition and face detection to human computer interface design, driver behavior analysis, and compression techniques development. By locating the position of the eyes, the gaze can be determined.

A large number of works have been published in the last decade on this subject. Generally the detection of eyes is done in two steps: locating face to extract eye regions and then eye detection from eye window. The face detection problem has been faced up with different approaches: neural network, principal components, independent components, and skin color based methods. Recently, methods based on boosting have become the focus of active research. The eye detection is done in the face regions which have been already located [1], [2], [3].

Little research has been done, however, on the direct search for eyes in whole images. Some approaches are based on active techniques: they exploit the spectral properties of pupil under near IR illumination. For example, in [4] two near infrared multiple light sources synchronized with the camera frame rate have been used to generate bright and dark pupil images. Pupils can be detected by using a simple threshold on the difference between the dark and the bright pupil images. In [5], iris geometrical information is used for determining a region candidate that contains an eye in the whole image, and then the symmetry is used for selecting the pair of eyes. Although the detection rate is high, the assumption that the distance between the camera and the person does not change greatly limits its practical applicability.

The main objective of this work is to propose an eye pair detection algorithm that is applicable with a standard camera, in a real world condition, while skipping the initial segmentation step to extract the face region as commonly done in literature. The developed eye detection algorithm works on the whole grey-level image. The structure of the eye region is used as a robust cue to locate eye pair candidates. Eye pairs are extracted by using binary template matching and a

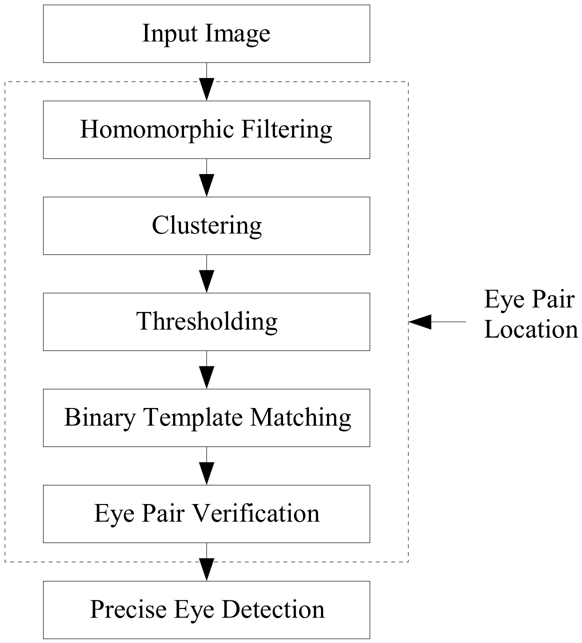


Fig. 1: Proposed eye detection method.

support vector machine (SVM). After locating eye pairs, an eye variance filter is used to detect two eyes precisely in the eye pair region.

2. Eye Pair Location

Before two eyes are detected precisely, an eye pair is located. The structure of the eye region is considered as a stable and robust feature which can help distinguish eye pair from other patterns. The proposed method uses this robust cue to locate eye pairs. The basic flowchart of the method is shown in Fig.1. Initially, the facial image is enhanced and is binaried to obtain the structure image. The structure image is defined as the binary image which contains the structure feature of human face. Then, a binary eye pair template is used to find eye pair candidates in the image. All the eye pair candidates are then rescaled to a fixed size and sent to an SVM classifier that verifies the candidates and obtains real eye pairs. Finally, eye pairs are located according to the verification results.

2.1 Preprocessing

The facial images processed in the template matching step are binary images that contain facial structure information. In order to obtain appropriately-segmented binary images, an image preprocessing is applied. To compensate for illumination variations and to obtain more image details, a homomorphic filter is used to enhance the brightness and the contrast of the images. Then a clustering algorithm is used to separate the facial feature from the skin. Binaried images are obtained through thresholding.

2.1.1 Homomorphic Filtering

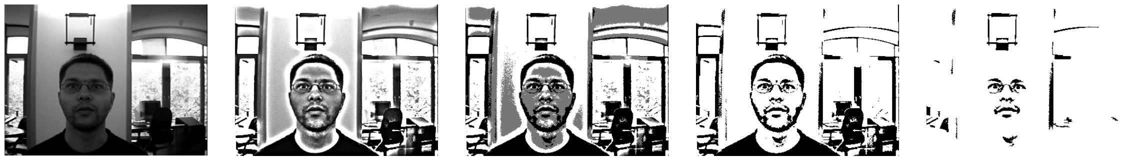
Homomorphic filtering [6] is a generalized technique for nonlinear image enhancement and correction. It simultaneously normalizes the brightness across an image and increases contrast.

An image can be expressed as the product of illumination and reflectance:

$$f(x,y) = i(x,y) \cdot r(x,y) \quad (1)$$

When the illumination is uniform, $i(x,y)$ is considered to be a constant, and the image is considered to be the reflectance of the object. However, the lighting condition is usually unequal. The illumination component tends to vary slowly, and, therefore, is represented by the lower frequency components in the frequency domain. The reflectance component, on the other hand, tends to vary rapidly and is represented by the higher frequency components in the frequency domain. If the illumination and reflectance can be acted upon separately, the illumination problem will be solved, and the image will be enhanced. Hence, the log transform is used to equation (1):

$$ln\, f(x,y) = ln\, i(x,y) + ln\, r(x,y) \quad (2)$$



(a) (b) (c) (d) (e)

Fig. 2: An example of preprocessing. (a) Original image. (b) Enhanced image. (c) Clustered image. (d) Binaried image. (e) Structure image.

Then the Fourier transform is applied to equation (2) and filtering is done in the frequency domain. The basic homomorphic filtering procedure is as follows:

$$f(x,y) \Rightarrow ln \Rightarrow FFT \Rightarrow H(u,v) \Rightarrow (FFT)^{-1} \Rightarrow exp \Rightarrow g(x,y) .$$

The illumination and reflectance turn to additive though log transform. Then, 2-D Fourier transform is used. The coordinate variables become u and v. $H(u,v)$ is the homomorphic filter function applied to the illumination and reflectance. After taking the inverse Fourier transform and the exponent transform, an enhanced image $g(x,y)$ is obtained. $H(u,v)$ used in this paper has the following form:

$$H(u,v) = (H_H - H_L)\cdot(1 - exp(-C \cdot \frac{D(u,v)}{D_0})) + H_L , \tag{3}$$

where $D(u,v)$ is the distance between point (u,v) and point of origin in frequency domain; D_0 is the threshold; C is sharpen parameter. If the parameters H_L and H_H are chosen to be $H_L < 1$ and $H_H > 1$, then the filter $H(u,v)$ will decrease the contribution of the low frequency (illumination) and amplify the contribution of the mid- and high frequencies (reflectance). As shown in Fig. 2 (a), the input image has low contrast due to illumination; segmentation results, therefore, are unlikely to be good. Fig. 2 (b) demonstrates the image enhanced by homomorphic filtering; the contrast is improved, and the details in the face region are enhanced.

2.1.2 Clustering and Thresholding

The intensity in the eye region and other facial features is dark in a grey-level facial image. The image has been enhanced through homomorphic filtering. Next, the features of interest are separated from the skin and other pixels by clustering the grey-level image into three clusters through the K-mean clustering algorithm. The lightest grey level that represents the light pixels in the image is set to 255, the intermediate that represents the skin is set to 128, and the darkest that represents both the features and other dark pixels of the image (for example the hair, beard and some dark background) is set to 0. Fig. 2 (c) shows the facial image processed by the clustering algorithm.

After clustering, a threshold is set to 128, so that only dark pixels remain, including eye pair structure. Then, a binary image is obtained, which obviously contains the facial structure. Fig. 2 (d) shows the thresholding result. Taking into account that the nonface area can influence the speed and the results of template matching, the oversize black area, which is useless in the binary image, is eliminated by the conventional connected components labeling process. Then the final feature image is obtained, as shown in Fig. 2 (e).

2.2 Binary Template Matching

In order to determine the set of rows that contain the eyes (eyes band), a binary template matching is applied to the feature image, searching for the two eyes [7]. The difficulty is that we

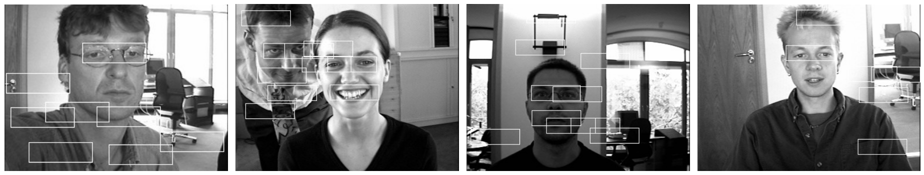


Fig. 3: Binary template matching results.

are not looking for an object with a fixed shape. For this reason a binary template is adopted that models two eyes in a very rough way. A single template has been used for all the images which are of significant different size, thus showing a desirable scale independence property.

Among the positions with the high cross correlation, all the candidates are extracted. In the proposed method, some in depth rotation of the face depth or rotation on-the-plane of the image are permitted as long as both eyes of the face are visible. Fig. 3 shows some matching results.

In order to detect faces in different scales, the input image is repeatedly scaled by a factor of 1.2 for 3 times. Faces that are the too small are ignored because the corresponding eye regions are also too small to precisely locate the eyes and detect their state. In each scale, all eye pair candidates are marked and verified by the eye verifier which will be described in the next section.

2.3 Eye Verifier

For the purpose of getting the successful face detection rate, the proposed method is followed by a simple eye verifier. Eye verifier is used instead of face verifier as proposed in e.g. [8, 9] because there may be occlusion on the region below eyes, and the exaggerated face expression appears below eyes which will also influence the results.

2.3.1 Support Vector Machines (SVM)

In this paper, SVM is selected as the classifying function. One distinctive advantage this type of classifier has over traditional neural networks is that SVMs can achieve better generalization performance [10]. Support vector machine is a pattern classification algorithm developed by Vapnik [11]. It is a binary classification method that finds the optimal linear decision surface based on the concept of structural risk minimization. As shown by Vapnik, this maximal margin decision boundary can achieve optimal worst-case generalization performance. Note that SVMs are originally designed to solve problems where data can be separated by a linear decision boundary. By using kernel functions (e.g. [12]), SVMs can be used effectively to deal with problems that are not linearly separable in the original space. Some of the commonly used kernels include Gaussian Radial Basis Functions (RBFs), polynomial functions, and sigmoid polynomials whose decision surfaces are known to have good approximation properties. Relying on the fact that the training data set is not linearly separable, a Gaussian Radial Basis Function (RBF) kernel is selected in this paper. The RBF kernel performs usually better for the reason that it has better boundary response as it allows for extrapolation. Our implementation of SVM is based on the SVM Matlab toolbox by A. Rakotomamonjy and S. Canu [13].

2.3.2 Eye Pair Candidates Verification

All the eye pair candidates (grey-level images) are extracted according to the results of binary template matching. Then they are normalized into the size of 25×8 pixels and verified by using the SVM to obtain real eye pairs.

The training data used for generating an eye verification SVM consists of 400 images of each class (eye-pair and non-eye-pair). Non-eye-pair images are much more diverse than eye pair images, so it is difficult to select standard non-eye-pair images for the training set. Selection of proper non-eye-pair images is very important for training the SVM because its performance is influenced by what kind of non-eye-pair images are used. In the initial stage of SVM training, non-eye-pair images are used that are similar to eye pair such as eyebrows, nostrils and other eye-pair-like patches. Non-eye-pair images are generated using the bootstrapping method proposed by Sung and Poggio [14]. For the RBF kernel parameter "gamma" set to 3 the trained SVM had 198 support vectors.

After the real eye pairs are obtained, the eye regions and the corresponding face regions are located according to the position of each eye pair.

3. Precise Eye Detection

After locating the eye pair, the precise eye detection is performed by using the eye variance filter method proposed by Feng and Yuen [15]. The change of grey intensity in the eye region containing white, pupil and eyelids is obvious on the human face. The variance on a domain is the second-order moment which indicates the measurement of variation of grey intensity. Based on these observations, an eye variance filter is developed. Applying the eye variance filter to an eye region, an strong response in the eye window is obtained while the response in a non-eye window is relatively low. The primary objective of using the variance filter is to precisely extract the two eyes in the eye region. The eye pair image is the enhanced image processed by homomorphic filter and extracted from the facial image. The left eye and the right eye are detected separately by dividing the eye region into two parts.

3.1 Eye Filter Construction

Let $I(x,y)$ be an eye image, the variance on a domain Ω is defined as

$$\sigma_\Omega^2 = \frac{1}{A_\Omega} \sum_{(x,y)\in\Omega} [I(x,y) - \overline{I_\Omega}]^2 , \tag{4}$$

where A_Ω and $\overline{I_\Omega}$ represent the area and the average grey intensity on the domain Ω, respectively. From the definition, the statistics σ_Ω^2 has two properties. First, σ_Ω^2 is rotation invariant in domain Ω. Second, σ_Ω^2 reflects grey intensity variations rather than the exact shape on the domain.

For an image $I(x,y)$, its variance image is defined as follows:

$$I_\sigma(i,j) = \sigma_{\Omega_{ij}} , \ \Omega_{ij} = \{(i-1)l + 1 \le x < il, \ (j-1)j + 1 \le y < jl\} . \tag{5}$$

To construct an eye variance filter, 30 eye images of size 21×15 with different states from different persons are selected as a training image set. The size is set according to the size of eye pair template which is used to find eye pair candidates. Fig. 4 shows these eye images. For each image, a 3×3 non-overlapped subblock is selected to calculate the variance. The corresponding variance images are determined and shown in Fig. 5. The eye variance filter F is constructed by calculating the average of all the variance images

$$F = \frac{1}{N} \sum_{i=1}^{N} I_\sigma^i . \tag{6}$$

The eye variance filter is shown in Fig. 6(a) and its 3D plot is shown in Fig. 6(b).

Fig. 4: Eye images.

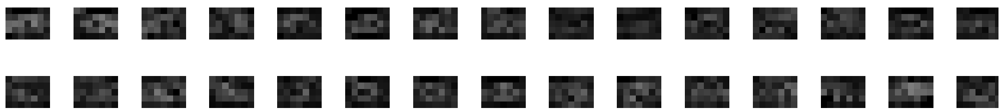


Fig. 5: Variance eye images.

3.2 Evaluation of the Eye Filter

In order to detect eyes using an eye variance filter, the related coefficient is calculated between the filter and each variance image block. The related coefficient is defined as follows:

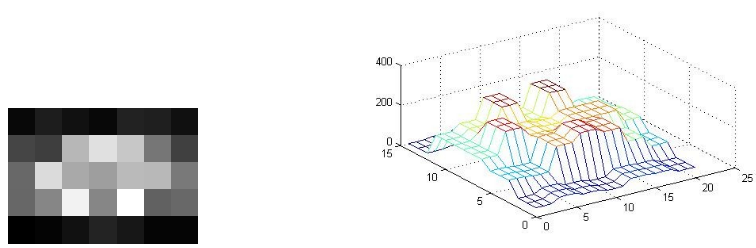
$$R(I_{\sigma_i}, F) = \frac{E[(\xi_{I_{\sigma_i}} - E(\xi_{I_{\sigma_i}}))(\xi_F - E(\xi_F))]}{\sqrt{D(\xi_{I_{\sigma_i}})D(\xi_F)}}, \qquad (7)$$

where ξ_{σ_i} and ξ_F are the concatenated vectors of the variance image I_{σ_i} and F, respectively; I_{σ_i} is an image block obtained by moving a scan window on the eye image; $E(\cdot)$ and $D(\cdot)$ represent the mathematical expectation and variance of the random variable. The related coefficient between the eye variance filter and each image block is calculated. The maximum related coefficient is then compared with the threshold determined by all of the related coefficients obtained in the experiment. If the maximum related coefficient is greater than the threshold, the eye is detected. If the maximum related coefficient is less than the threshold, the detection fails. The proposed eye variance filter can extract two eyes precisely if the change of grey intensity on eye region is normal.

To detect two eyes in the eye region, the extracted eye pair image, shown in Fig. 7 (a), is transformed into a variance image, shown in Fig. 7 (b). The result of convolution of the eye variance filter with the variance image is shown in Fig. 7 (c). Fig. 7 (d) demonstrates the final detection result.

4. Experimental Results

The proposed method was tested on the BioID face database [16]. The BioID face database consists of 1521 images (384 × 286 pixels, grey level) of 23 different test persons and has been recorded during several sessions at different places. This set features a larger variety of illumination, background and face size. It stresses real world conditions. So it is believed to be more difficult than other datasets containing images with uniform illumination and background.



(a) (b)

Fig. 6: Eye variance filter. (a) Eye variance filter. (b) The 3D plot of eye variance filter.

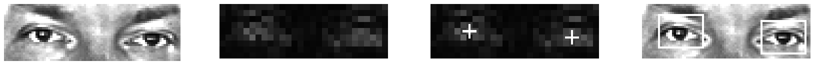


Fig. 7: Example of precise eye detection. (a) Eye pair image. (b) Variance eye pair image. (c) Response of the eye pair variance image. (d) Eye detection result.

The eye pair candidates can be selected successfully in most cases, no matter whether face patterns are in different scale, expression, and illumination conditions. The eye location is successful after eyes verification by using SVM in most case. The eye location rate is 95.6% in BioID database. The precise eye detection rate is 96.8%.

Typical results of eye detection with the proposed approach are demonstrated in Fig. 8. The input images vary greatly in background, scale, expression and illumination, the images also including partial face occlusions and glasses wearing. The eye pair and two eye windows which have been detected are marked with white rectangles.

5. Conclusion

In this paper, an efficient method for detecting eyes in still grey-level images with unconstrained background is presented . The structure of the eye region is used as a robust cue to find eye pairs. To obtain structure images, the preprocessing is applied to input images. Homomorphic filtering is applied to enhance the contrast of dark regions; therefore, facial images with poor contrast are enhanced, which makes the clustering process more effective. Eye pair candidates are extracted by using a binary template matching technique. However, this method can not deal with large out-plane face rotation because the structure of the eye region changes. In-plane face rotation can be solved by rotating facial images to predefined degrees. Precise eye windows are detected by using an eye variance filter on the located eye pair images.

The proposed method can deal with glasses wearing and partial face occlusions. However, the eye detection will fail if the reflection of glasses is too strong or eyes are closed. If the reflection of glasses is too strong, the eyes can not be extracted in the clustering step; consequently, the proposed method will fail finding eye pairs in the template matching step. Closed eyes will not influence the results of eye location; however, the precise eye detection will fail because the intensity variance on the eye region is changed. Therefore, the preprocessing algorithm needs to be improved.

Current theoretical analysis and implementation show that AdaBoost [17] provides a good framework to select features and combine classifiers. The future work will be focused on detecting faces and eyes by using automatically learned features.

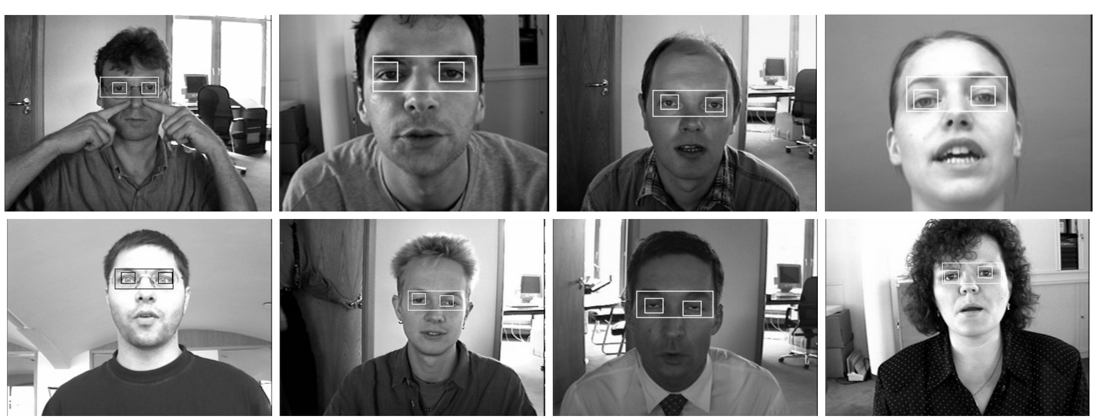


Fig. 8: Eye detection results.

Acknowledgments

This work is supported by the Project of Science and Technology Plan of Jiangsu Province under grant BG2005008, and NSF of China under grant 60473039, 60503026.

References

[1] T. Kawaguchi and M. Rizon, "Iris detection using intensity and edge information," *Pattern Recognition*, vol. 36, pp. 549–562, 2003.

[2] S. Baskan, M. Bulut, and V. Atalay, "Projection based method for segmentation of human face and its evaluation," *Pattern Recognition Letters*, vol. 23, pp.1623–1629, 2002.

[3] S. Sirohey, A. Rosenfiled, and Z. Duric, "A method of detection and tracking iris and eyelids in video," *Pattern Recognition*, vol. 35, pp. 1389–1401, 2002.

[4] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance," in *Proc. IEEE Conf.. Computer Vision and Pattern Recognition*, vol. 1, pp. 163–168, 2000.

[5] T. D'Orazio, M. Leo, G.Cicirelli, and A. Distante, "An algorithm for real time eye detection in face images," in *Proc. 17th Int. Conf. on Pattern Recognition*, vol. 3, pp. 278–281, 2004.

[6] A. V. Oppenheim, R. W. Schafer, and T. G. S. Jr, "Nonlinear filtering of multiplied and convolved signals," *IEEE Proc.*, vol. 56, no. 8, pp. 1264–1291, 1968.

[7] P. Campadelli and R. Lanzarotti, "Localization of facial features and fiducial points," in *Proc. Int. Conf. Visualization, Imaging and Image Processing*, pp. 491–495, 2002.

[8] J. X. Wu and Z. H. Zhou, "Efficient face candidates selector for face detection," *Pattern Recognition*, vol. 36, pp. 1175–1186, 2003.

[9] C. C. Han, H. Y. M. Liao, G. J. Yu, and L. H. Chen, "Fast face detection via morphology-based pre-processing," in *Proc. 9th Int. Conf. Image Analysis and Processing*, pp. 469–476, 1998.

[10] C. A. Waring and X. W. Liu. "Face detection using spectral histograms and SVMs," *IEEE Trans. Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 35, no. 3, pp. 467–476, 2005.

[11] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[12] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 130–136, 1997.

[13] S. Canu, Y. Grandvalet, V. Guigue, and A. Rakotomamonjy. (2005). SVM and kernel methods Matlab toolbox. *Perception Systèmes et Information, INSA de Rouen, Rouen, France* [Online]. Available: http://asi.insa-rouen.fr/~arakotom/toolbox/index.html

[14] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20. no. 1, pp. 39–51, 1998.

[15] G. C Feng and P. C Yuen, "Multi-cues eye detection on grey intensity image," *Pattern Recognition*, vol. 34, no. 5, pp. 1033–1046, 2001.

[16] O. Jesorsky, K. Kirchberg, and R. Frischholz. (2001). The BioID face database. *HumanScan* [Online]. Available: http://www.bioid.com/downloads/facedb/index.php

[17] Y. Freund, R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *In Computational Learning Theory: Eurocolt 95*, Springer-Verlag, pp. 23–37, 1995.