Eye Detection using Wavelets and ANN

Rakhi C. Motwani
University of Nevada, Reno
Dept of Comp. Sci. and Engr.
University of Nevada, Reno
Reno, NV 89557 USA
(775) 448-7672
rakhi@cs.unr.edu

Mukesh C. Motwani
Image Process Technology, Inc.
1776 Back Country Road
Reno, NV 89521
USA
(775) 853-7897
mukesh@image-process.com

Dr. Frederick C. Harris, Jr.
University of Nevada, Reno
Dept of Comp. Sci. & Engr.
University of Nevada, Reno
Reno, NV 89557 USA
(775) 784-6571
fredh@cs.unr.edu

Abstract

In this paper we suggest a novel method which is robust and efficient in extracting eye windows using Wavelets and Neural Networks. Wavelet analysis is used as a pre-processor for a back propagation neural network with conjugate gradient learning. The inputs to the neural network are the wavelet maxima neighborhood coefficients of face images at a particular scale. The output of the neural network is the classification of the input into an eye or non-eye region. An accuracy of 88% is observed for test images under different environment conditions not included during training.

1. Introduction

Recognition of human faces out of still images or image sequences is an actively developing research field. There are many different applications for systems coping with the problem of face localization and recognition e.g. model based video coding, face identification for security systems, gaze detection and human computer interaction. The detection and location of the face as well as the extraction of facial features from the images are crucial. Due to variations in illumination, background, visual angle and facial expressions, the problem is complex. In the first step of face recognition, the localization of facial regions within the facial contours is followed by the detection of facial features such as eyes, nose and mouth.

In this paper, we describe a novel algorithm for eye detection that is robust against changes in light conditions, visual angle, and noise in addition to being low in computational cost.

Many factors conspire to considerably complicate images of actual eyes: glasses, eyebrows, eyelashes, makeup, perspiration, the wide range of deformation and movement of the eyelid, different directions of gaze, and variations in lighting and viewpoint. Of even greater concern is the fact that the structure of an image of an eye varies considerably with distance and camera resolution. Even at the coarsest resolution the most significant structure is an intensity valley caused by the self-shadowing of the eye socket. One of the motivations of our work is to detect facial features for 3D modeling of the human head, which is used to estimate pose for the human-machine interaction. The following section summarizes the various techniques that have been utilized thus far in the filed of eye detection research.

2. Related Work

A lot of research work has been published in the field of eye detection in the last decade. Various techniques have been proposed using texture, depth, shape and color information or combinations of these for eye detection. Vezhnevets et al. [1] focus on several landmark points (eye corners, iris border points), from which the approximate eyelid contours are estimated. The iris center and radius is detected by looking for a circle separating dark iris and bright sclera. The upper eyelid points are found using on the observation that eye border pixels are significantly darker than surrounding skin and sclera. The detected eye boundary points are filtered to remove outliers and a polynomial curve is fitted to the remaining boundary points. The lower lid is estimated from the known iris and eye corners. Reinders et al. [2] present a method where based on the technique of template matching the positions of the eyes on the face image can be followed throughout a sequence of video images. To increase the robustness of the tracking scheme the method automatically generates a codebook of images representing the encountered different appearances of the eyes. Yuille et al. [3] first proposed using deformable templates in locating human eye. The weaknesses of the deformable templates are that the processing time is lengthy and success relies on the initial position of the template. Lam et al. [4] introduced the concept of eye corners to improve the deformable template approach. Saber et al. [5] and Jeng et al. [6] proposed to use facial features geometrical structure to estimate the location of eyes. Takacs et al. [7] developed iconic filter banks for detecting facial landmarks. Projection functions have also been employed to locate eye windows [9, 10, 11, 12]. Feng and Yeun [9] developed a variance projection function for locating the corner points of the eye. Zhou and Geng [10] propose a hybrid projection function to locate the eyes. By combining an integral projection function, which considers mean of intensity, and a variance projection function, which considers the variance of intensity, the hybrid function better captures the vertical variation in intensity of the eyes. Kumar et al. [11] suggest a technique in which possible eye areas are localized using a simple thresholding in color space followed by a connected component analysis to quantify spatially connected regions and further reduce the search space to determine the contending eye pair windows. Finally the mean and variance projection functions are utilized in each eye pair window to validate the presence of the eye. Feng and Yeun [12] employ multi cues for eye detection on gray images using variance projection function

The most common approach employed to achieve eye detection in real-time [13, 14, 15, 16] is by using infrared lighting to capture the physiological properties of eyes and an appearance-based model to represent the eye patterns. The appearance-based approach detects eyes based on the intensity distribution of the eyes by exploiting the differences in appearance of eyes from the rest of the face. This method requires a significant number of training data to enumerate all possible appearances of eyes i.e. representing the eyes of different subjects, under different face orientations, and different illumination conditions. The collected data is used to train a classifier such as a neural net or support vector machine to achieve detection.

Various other methods that have been adopted for eye detection include wavelets, principal component analysis, fuzzy logic, support vector machines, neural networks, evolutionary computation and hidden markov models. Huang and Wechsler [17] perform the task of eye detection by using optimal wavelet packets for eye representation and radial basis functions for subsequent classification of facial areas into eye and non-eye regions. Filters based on Gabor wavelets to detect eyes in gray level images are used in [18]. Talmi et al. and Pentland et al. [19, 21] use principal component analysis to describe and represent the general characteristics of human eyes with only very few dimensions. In [19] eigeneyes are calculated by applying Karhunen-Loeve-Transformation to represent the major characteristics of human eyes and are stored as reference patterns for the localization of human eyes in video images. Given a new input image of the same size, it is projected into the eigeneye space. The produced vector describes the similarity of this new image to the eigeneyes. If similarity measure (the Euclidean distance between the mean adjusted input image and its projection onto the eigeneye space) smaller than a threshold, the new image is classified as an eye region. Hjelms and Wroldsen [20] utilize Gabor Filters and PCA for eye detection.

Li et al. [22] construct a fuzzy template which is based on the piecewise boundary. A judgment of eye or non-eye is made according to the similarity between the input image and eye template. In the template, the eyelid is constructed by a region of adjacent segments along the piecewise boundary. Each segment in the fuzzy template is filled with the darkest intensity value within this segment. This makes the method have invariance to slight change of eye images in size and shape and it gives the method high robustness to different illumination and resolution for input images, by increasing the contrast between eyelid region and its adjacent regions. Furthermore, histogram equalization and multi-thresholding image binarization measures are also taken to deal with the problem of different illumination, contrast and resolution for input images. Huang et al. [23] employ SVM classifier to scan for possible eye locations across the whole face.

Genetic algorithms have been exploited for eye detection [24] by fitting image distributions (mean, entropy and standard deviation) between training and probed image. Using evolutionary computation based methods the most likely eye locations are discovered and a model-based eye recognition component probes the suggested locations for actual eye detection. Bala et al. [25] utilize genetic algorithms for feature selection. Base features representations and visual routines for eye detection represented as decision trees are evolved. The proposed technique uses induction of decision trees for the evaluation function, in hybrid architecture. Koch et al. [26] use a neural network to scan an input window of pixels across the image, where each gray value in the input window serves as an input for the neural network. The neural network is then trained to give a high response when the input window is centered on the eye. After scanning the entire image, the position with the highest response then reveals the center position of the eye in the image. In order to allow applicability in far more general situations, i.e. covering a large variety of eye appearances, allowing rotation and scaling, and allowing different lighting conditions the eyes are located by locating micro-features, rather than entire eyes and the neural network responses are post-processed by a probabilistic method which exploits the geometrical information about the micro-features.

Samaria and Harter [27] employ stochastic modeling, using Hidden Markov Models (HMMs) to holistically encode feature information. When frontal images of faces are sampled using top-bottom scanning, there is a natural order in which the features appear and this is conveniently modeled using a top-bottom HMM. The HMM lead to the efficient detection of eye strips.

3. Proposed Method

The system consists mainly of two stages – training and detection stage. A block diagram of these two stages is shown in Figure 1.

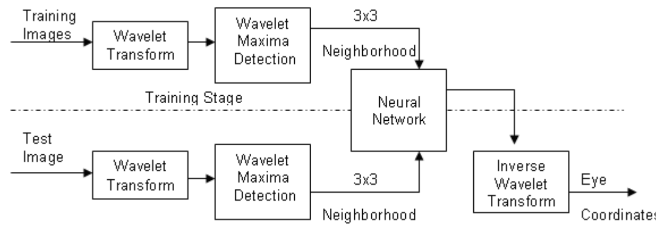

Figure 1 - Block Diagram

3.1 Acquisition Of Training Data

The training data typically consists of 50 images of different persons with different hair styles, different illumination conditions and varying facial expressions. Some of the images have different states of the eye such as eyes closed. The size of the images varies from 64x64 to 256x256.

3.2 Discrete Wavelet Transform

Wavelet decomposition provides local information in both space domain and frequency domain. Despite the equal subband sizes, different subbands carry different amounts of information. The letter ‘L’ stands for low frequency and the letter ‘H’ stands for high frequency. The left upper band is called LL band because it contains low frequency information in both the row and column directions. The LL band is a coarser approximation to the original image containing the overall information about the whole image. The LH subband is the result of applying the filter bank column wise and extracts the facial features very well. The HL subband, which is the result of applying the filter bank row wise, extracts the outline of the face boundary very well. While the HH band shows the high frequency component of the image in non-horizontal, non-vertical directions it proved to be a redundant subband and was not considered having significant information about the face. This observation was made at all resolutions of the image. This is the first level decomposition. A CDF (2, 2) bi-orthogonal wavelet is used. Gabor Wavelets seem to be the most probable candidate for feature extraction. But they suffer from certain limitations i.e. they cannot be implemented using

Lifting Scheme and secondly the Gabor Wavelets form a non-orthogonal set thus making the computation of wavelet coefficients difficult and expensive. Special hardware is required to make the algorithm work in real time. Thus choosing a wavelet for eye detection depends on a lot of trial and error. Experiments were done with Haar and Daub4 Wavelets and no improvement in performance was observed. Discrete Wavelet Transform is recursively applied to all the images in the training data set until the lowest frequency subband is of size 32x32 pixels i.e. the LH subband at a particular level or depth of DWT is of size 32x32. The original image and it's wavelet decomposition is shown in Figure 2 and 3. The LH subband at resolution 32x32 is shown in Figure 4.

Figure 2 - Original Image

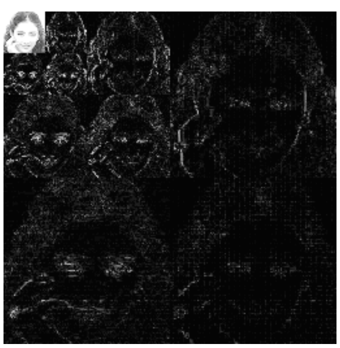

Figure 3 - Discrete Wavelet Transform of Original Image

We take the modulus of the wavelet coefficients in the LH subband. Experiments were performed to go to a resolution even coarser than 32x32. However, it was observed that in certain cases the features would be too close to each other and it was difficult even manually too to separate them. This would burden the Neural Network model and a small error in locating the eyes at this low resolution would result in a large error in locating the eyes in the original image.

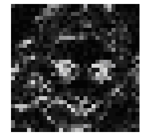

Figure 4 - LH subband of resolution 32x32

3.3 Detection of Wavelet Maxima

Our approach to eye detection is based on the observation that, in intensity images eyes differ from the rest of the face because of their low intensity. Even if the eyes are closed, the darkness of the eye sockets is sufficient to extract the eye regions. These intensity peaks are well captured by the wavelet coefficients. Thus, wavelet coefficients have a high value at the coordinates surrounding the eyes. We then detect the wavelet maxima or the wavelet peaks in this LH subband of resolution 32x32. Note that several such peaks are detected, which can be the potential locations of the eyes. The intensity peaks are shown in Figure 5 and 6.

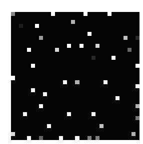

Figure 5 - Wavelet Maxima Detection

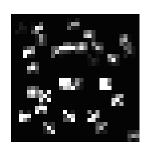

Figure 6 - Subband with Peaks replaced by 3x3 neighborhood wavelet coefficients from the previous image

3.4 Neural Network Training

The wavelet peaks detected are the center of potential eye windows. We then feed 3x3 neighborhood wavelet coefficients of each of these local maxima's in 32x32 LH subbands of all training images to a Neural Network for training. The Neural Network has 9 input nodes, 4 hidden nodes, and 2 output nodes. A diagram of the Neural Network architecture is shown in Figure 7. A (0, 1) at the output of Neural Network indicates an eye at the location of the wavelet maxima whereas (1, 0) indicates a non-eye. Two output nodes instead of one were taken to improve the performance of the Neural Network. MATLAB's Neural Network Toolbox was used for simulation of the back propagation Neural Network. A conjugate gradient learning rate of 0.7 was chosen while training. This completes the training stage.

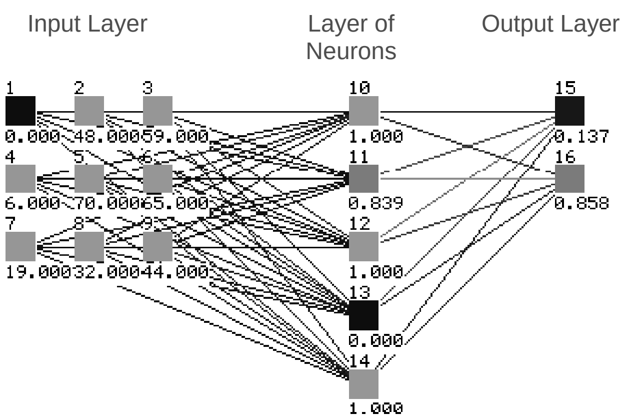

Figure 7 - Neural Network Architecture

3.5 Location of Eye Coordinates In Test Image

Discrete Wavelet Transform is recursively applied to the test image until the lowest frequency subband is of size 32x32 pixels i.e. the LH subband is of size 32x32. The test image size maybe an integer multiple of 64x64. The absolute values of the wavelet coefficients are taken in this subband. Wavelet peaks in this subband are detected which are location of the potential eye windows. These peaks are then replaced with 3x3 neighborhood wavelet coefficients from the previous image and then fed to the Neural Network. The Neural Network classifies each of the peaks as eye or non-eye in this 32x32 subband.

Thus we need to post process the image to locate the coordinates of the eyes in the image of original size. If post processing is not done the algorithm can only be used to indicate the presence of eyes in the image of original size. Post processing is necessary for the algorithm to indicate the presence of eyes in the original image. The Embedded Zero Tree (EZW) coder [29] exploits the relations between wavelet coefficients in different subbands. Figure 8 depicts the relations between wavelet coefficients in different subbands.

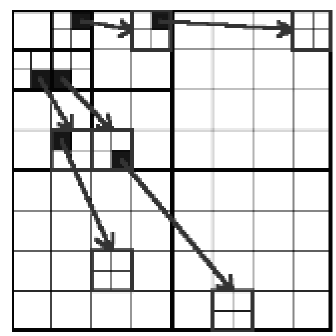

Figure 8 - The relations between wavelet coefficients in different subbands

The EZW encoder is based on two important observations – i) a coefficient in a low subband can be thought of as having descendants in the next higher subband and ii) large wavelet coefficients are more important than small wavelet coefficients. EZW encoder uses the dependency between the wavelet coefficients across different scales to efficiently encode large parts of the image which are below a certain threshold. Thus, if we change the value of a larger wavelet coefficient and then take the inverse DWT of the image, the change will be reflected in the reconstructed image in the region surrounding the spatial location of the modified wavelet coefficient. This logical reasoning was verified experimentally and it was observed that if the wavelet coefficients corresponding to the location of the eyes in the LH subband of resolution 32x32 are reset to a very large negative value and then the Inverse is Discrete Wavelet Transform of the image taken, a petal shaped distortion occurred at the location

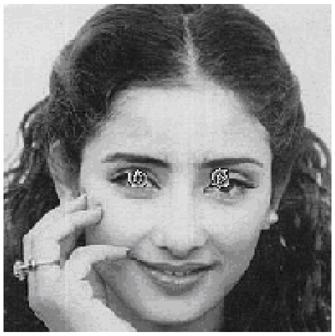

Figure 9 - Eyes detected

of the eyes in the original image size, as shown in Figure 9. This image can be subtracted with the original image to replace the petal shaped distortion with an eye window.

4. Performance

A number of experiments were done to test the robustness of the algorithm and to increase the accuracy of eye detection. Various architectures of Neural Networks with different learning rates were tried and it was found that back propagation with conjugate gradient learning seemed to be the best choice. A very high learning rate of 0.7 was chosen because the learning algorithm was getting trapped in local minima while training the network. Final training was stopped when the error graph, as shown in Figure 10, didn't show any significant fluctuation.

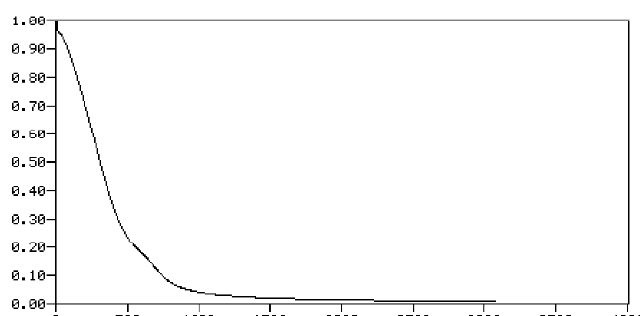

Figure 10 - Conjugate Training Error Curve

An experiment was done in which the face was analyzed using wavelet packets and it was found that most of the information was retained by the low frequency subbands and the high frequency packets had no information. Images with different states of the eye (closed, open, half open, looking sideways, head tilted etc.) and varying eye width were chosen. The eye positions found were compared with the positions that were pointed out manually. The eyes were correctly located when its location is within two pixels, in both x and y directions, of the manually assigned point. The variation of 2 pixels is deliberately allowed, to compensate for the inaccuracies in the location of eyes during training. An accuracy of 88% was observed in the final location of the eyes. A database of 75 test images was evaluated for performance. Most of the images were from the Olivetti Research Laboratory in Cambridge, UK. All these test images were captured in totally different environment conditions and were not included while training the Neural Network. Most of the error cases occurred in images with complex background. Also there was an error in accurately determining the exact location of the eyes since a 1 pixel shift at a resolution of 32x32 corresponded to a larger shift in the exact location of the eye. In some cases the Neural Network classified only 1 peak as an eye in spite of the presence of 2 eyes in the image. In a few cases observations were made in which regions of the face not belonging to the eyes were detected as eyes. In other cases more than 2 eyes were indicated in the image. Face images with closed eyes were also tested and the eyes were located as shown in Figure 11.

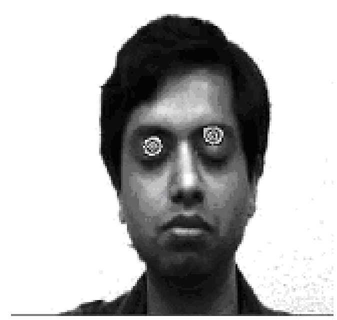

Figure 11 - Person with closed eyes

In contrast, the performance of this algorithm which uses wavelets as a preprocessor to Neural Networks, the algorithm with only Neural Networks achieved an accuracy of 61% in detecting the exact location of the eyes.

5. Conclusion

This paper gives a new dimension to the existing eye detection algorithms. The present algorithm is robust and at par with the other existing methods but still has a lot of scope for improvement. This paper proposed a wavelet subband approach in using Neural Networks for eye detection. Wavelet Transform is adopted to decompose an image into different subbands with different frequency components. A low frequency subband is selected for feature extraction. The proposed method is robust against illumination, background, facial expression changes and also works for images of different sizes. However, a combination of information in different frequency bands at different scales, or using multiple cues can even give better performance. Further studies in using Fuzzy Logic for data fusion of multiple cues will be our future direction.

6. References

[1] Vezhnevets V., Degtiareva A., "Robust and Accurate Eye Contour Extraction", *Proc. Graphicon*-2003, pp. 81-84, Moscow, Russia, September 2003.

[2] M.J.T. Reinders, "Eye tracking by template matching using an automatic codebook generation scheme", *Third annual conference of the Advanced School for Computing and Imaging*, pp. 85-91, Heijen, The Netherlands, June 1997.

[3] A. L. Yuille, P. W. Hallinan, D. S. Cohen, ”Feature extraction from faces using deformable templates”, *International Journal of Computer Vision* 8(2) (1992) 99-111.

[4] K. M. Lam, H. Yan, ” Locating and extracting the eye in human face images”, *Pattern Recognition,* Vol. 29, No. 5 (1996) 771-779.

[5] E. Saber and A.M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions", *Pattern Recognition Letters*, Vol. 19, No. 8, pp. 669--680, 1998.

[6] Jeng, S.H., Liao, H.Y.M., Han, C.C., Chern, M.Y., Liu, Y.T., “Facial Feature Detection Using Geometrical Face Model: An Efficient Approach”, *Pattern Recognition*, Vol. 31, No. 3, March 1998, pp. 273-282.

[7] Takacs, B., Wechsler, H., "Detection of faces and facial landmarks using iconic filter banks", *Pattern Recognition,* Vol. 30, No. 10, October 1997, pp. 1623-1636.

[8] L. Zhang, "Estimation of eye and mouth corner point positions in a knowledge-based coding system", *SPIE Vol. 2952, Digital Compression Technologies and Systems for Video Communications*, Berlin, October 1996, pp. 21-28.

[9] Feng, G.C., Yuen, P.C., “Variance Projection Function and its Application to Eye Detection for Human Face Recognition”, *Pattern Recognition Letters,* Vol. 19, July 1998, pp. 899-906.

[10] Z.-H. Zhou and X. Geng, "Projection Functions for Eye Detection", *Pattern Recognition*, 2004, 37(5): 1049-1056.

[11] Kumar, Thilak R and Raja, Kumar S and Ramakrishnan, “Eye detection using color cues and projection functions”, *Proceedings 2002 International Conference on Image Processing ICIP*, pages Vol.3 337-340, Rochester, New York, USA.

[12] Guo Can Feng and Pong C. Yuen, "Multi-cues eye detection on gray intensity image", *Pattern Recognition.* Vol 34, 2001, 1033-1046.

[13] X. Liu, F. Xu and K. Fujimura, “Real-Time Eye Detection and Tracking for Driver Observation Under Various Light Conditions”, *IEEE Intelligent Vehicle Symposium*, Versailles, France, June 18-20, 2002

[14] Zhiwei Zhu, Kikuo Fujimura, Qiang Ji, “Real-Time Eye Detection and Tracking Under Various Light Conditions and Face Orientations”, *ACM*

SIGCHI Symposium on Eye Tracking Research & Applications, March 25th-27th 2002, New Orleans, LA, USA.

[15] A. Haro, M. Flickner, and I. Essa, "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance", *Proceedings IEEE CVPR 2000*, Hilton Head Island, South Carolina, June 2000.

[16] C. Morimoto, D. Koons, A. Amir, and M. Flickner, "Real-Time Detection of Eyes and Faces", *Proceedings of 1998 Workshop on Perceptual User Interfaces*, pages 117-120, San Francisco, CA, November 1998.

[17] J. Huang and H. Wechsler, "Eye Detection Using Optimal Wavelet Packets and Radial Basis Functions (RBFs)", International *Journal of Pattern Recognition and Artificial Intelligence,* Vol 13 No 7, 1999

[18] S. A. Sirohey and Azriel Rosenfeld, "Eye detection in a face image using linear and nonlinear filters", *Pattern Recognition.* Vol 34. 2001, 1367-1391.

[19] K. Talmi, J. Liu, "Eye and Gaze Tracking for Visually Controlled Interactive Stereoscopic Displays", *Signal Processing: Image Communication* 14 (1999) 799-810 Berlin, Germany

[20] Erik Hjelms and Jrn Wroldsen, "Recognizing faces from the eyes only", *Proceedings of the 11th Scandinavian Conference on Image Analysis*, 1999.

[21] A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition", *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 1994, pp.84-91.

[22] Ying Li, Xiang-lin Qi and Yun-jiu Wang, "Eye detection using fuzzy template matching and feature-parameter-based judgement", *Pattern Recognition Letters*, Volume 22, Issue 10, pp. 1111-1124, Aug. 2001.

[23] Huang, J., D. Ii, X. Shao, and H. Wechsler, "Pose Discrimination and Eye Detection Using Support Vector Machines (SVMs)", *Proceeding of NATO-ASI on Face Recognition: From Theory to Applications,* Springer Verlag, pp.528-536.

[24] Huang, J., C. Liu, and H. Wechsler, (1998). "Eye Detection and Face Recognition Using Evolutionary Computation", *Proceeding of NATO-ASI on Face Recognition: From Theory to Applications*, Springer Verlag, pp.348-377.

[25] Bala, J., Huang, H., Vafaie, H., Wechsler, H., "Visual Routine for Eye Detection Using Hybrid Genetic Architectures", *Proc. ICPR'96*, Vienna, Austria, Vol. III, pp. 606-610, 1996

[26] M.J.T. Reinders, R.W.C. Koch, and J.J. Gerbrands, "Tracking facial features in image sequences using neural networks", *Proceedings of the Second international conference on Automatic Face and Gesture Recognition, 1996*

[27] F. S. Samaria, A. C. Harter, "Parameterization of a stochastic model for human face identification", *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, 1994, pp.138-142.

[28] W. Sweldens and P. Schrder, "Building your own wavelets at home", *Wavelets in Computer Graphics*, pp. 15--87, ACM SIGGRAPH Course notes, 1996.

[29] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients", *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, pp. 3445-3463, Dec. 1993.