Eye Detection by Haar wavelets and cascaded Support Vector Machine

Vishal Agrawal
B.Tech 4th Year
Guide: Simant Dubey / Amitabha Mukherjee
Dept of Computer Science and Engineering
IIT Kanpur - 208 016
INDIA

Contents

1 Title

Eye detection by Haar wavelet and Support Vector Machine.

2 Motivation

Accuracy of the face recognition system depends on the accurate localization of the facial features. These facial features either are used a classifier directly or for normalizing the images. Out of several of facial features eye is one of the most important facial feature, once the eye positions are detected then the other facial features can be detected very easily. Symmetry of the eyes make them extremely useful in face recognition application [1] [8]. Eye separation generally does not change significantly with facial expression, nor with up and down movement of the face, therefore eye separation distance is generally used for normalizing the face images. Face alignment is also determined using the positions of eye therefore an accurate eye localization algorithm is necessary for accurate face recognition.

Moreover eye detection is also very useful for gaze detection. Gaze determination has many applications: computer interfaces for helping the handicapped people, car driver's behavior understanding etc.

3 History

Eye being an important feature for face recognition algorithms, a lot of work has been done in the past decades but it is still an open problem. Here I will talk about some of the eye detection algorithms which have been proposed in recent years -

3.1 Robust Precise Eye localization by AdaBoost and SVM Techniques -

In this article [9] an approach for eye detection was described using AdaBoost and SVM techniques. This method uses a hierarchical cascade classifier based on AdaBoost statistical learning method combined with SVM classifier.

Training and detection of Hierarchical AdaBoost detector with SVM post classifier -

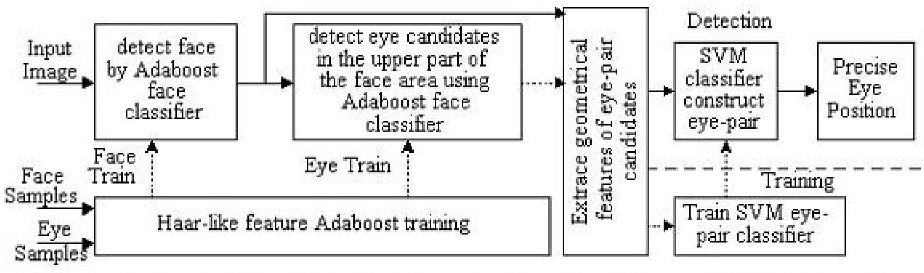

Figure from *X. Tang's Robust Precise Eye localization*[9]

First an AdaBoost face detector is applied to get the face on the whole image then an AdaBoost eye detector is applied to the upper part of the face to detect the possible eye candidates within the face area. The discard of the lower part of the image is based on the prior knowledge of the face appearance and it permits to reduce the computational complexity.Finally the precise eye positions are decided by the eye-pair SVM classifier which uses geometrical and relative position information of eye-pair and the face and those classified as admissible configurations are averaged together.

Due to geometrical constraints, the method is declared to work on images with head rotation up to 10^o both in and out of the plane.

3.2 Automatic Eye Detection and its Validation -

Due to its computational efficiency Haar wavelets are widely used in the field of image processing. But Haar wavelets have a little discriminating power, moreover features represented by Haar wavelets are square in shape. However, for eye detection the most distinguish feature is the pupil which has a round shape.

In this paper [10] the author propose to statistically learn discriminant features to characterize the eyes. Based on the distribution of discriminant features it learn probabilistic classifiers to sepearte eyes and non-eyes. Then multiple classifiers are combined in AdaBoost to form a robust and precise eye detector.

This method uses the RNDA (Recursive Non-parametric Discriminant Analysis) to extract more effective features, besides providing sample weights useful for the AdaBoost algorithm. The final classifier consists of two layers: the first has only two features thus very fast, while the second has about 100 features to refine the search.

This paper shows that this approach has more accuracy then Haar wavelets based approach.

3.3 Regression and Classification approaches to eye localization in face images -

This paper [4] investigates the three approaches to eye localization: a regression approach aiming to directly minimize errors in the predicted eye positions, a simple Bayesian model of eye and non-eye appearance, and a discriminative eye detector trained using AdaBoost.

All these methods are trained and tested on the same images and applied in cascade to a face detector module (Viola-Jones). Results show that simple Bayesian model outperforms the other approaches. Pointing out the difficulty in using classifier for eye detection.

3.4 Precise Eye localization through a general-to-specific model definition -

In this paper [2] the method described uses two SVM (Support Vector Machines) trained on properly selected Haar wavelet coefficients. This approach can be applied in cascade to any face detector that returns a rough estimation of the face position and scale.

The system is top-down and consists of two modules: the eye detector and the eye localizer. Eye detector is applied on the output of face detector and serves two purposes: it not only produces a rough localization of the general eye features, it also validates the output of the face detector. Then the eye localizer is applied in cascade to this in order to refine the localization precision by using the specific eye pattern definition.

Both the module is built using SVM trained in optimally selected Haar wavelet coefficients. The system proves to be robust to background clutter, to moderate illumination changes and to head rotations up to 20^o both in and out of the plane.

3.5 Feature detection and Tracking with Constrained Local Model -

This paper [3] describes an efficient and robust model matching method which uses a joint shape and texture appearance model to generate a set of region template detectors. The model is fitted to an unseen image in an iterative manner by generating templates using the joint model and the current parameter estimates, correlating the templates with the target image to generate response images and optimizing the shape parameters so as to maximize the sum of responses.

CLM Search Algorithm -

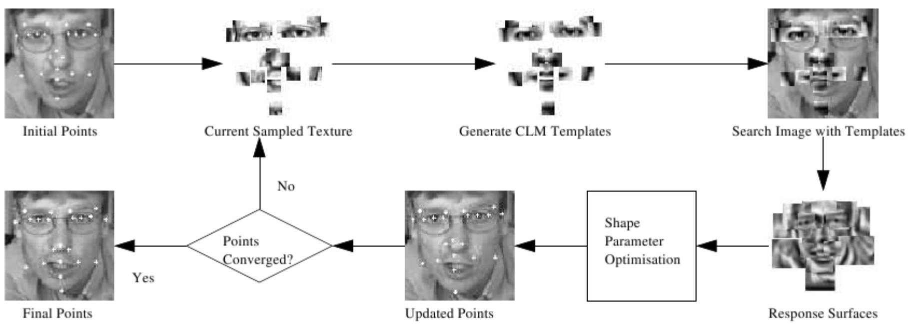

Figure from *D. Christinacce and T. Cootes's Feature detection and Tracking with CLM* [3]

Paper shows that this algorithm gives improved localization accuracy on two publicly available databases.

4 Comparison

The following graphs display the performance of Campadelli's eye detector (SVM-1), eye localizer (SVM-2), and when available the performance achieved by the methods described in the previous section -

The cumulative distributions of eye detection and localization over different databases -

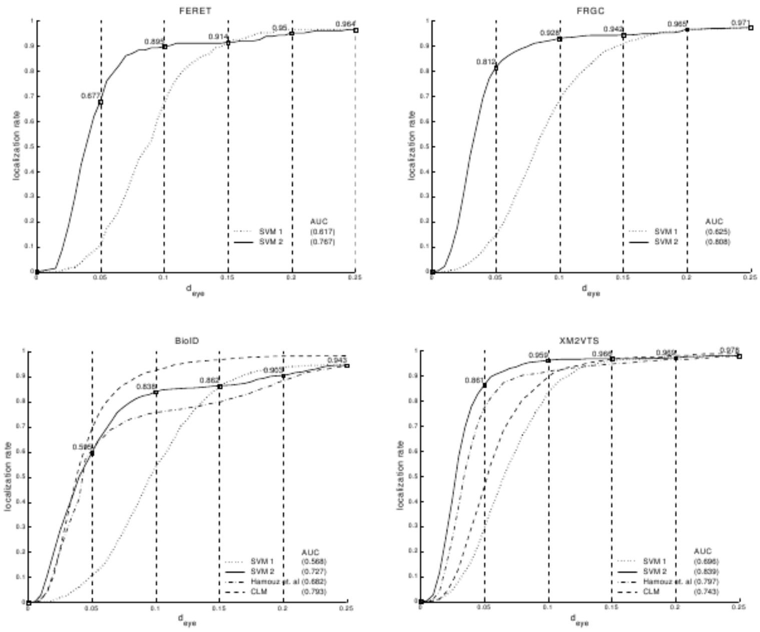

Figure from *C*ampadelli's Eye localization a survey[1]

As we can see Camdaelli's Precise eye localization method [2] shows good results on different databases. The CLM approach also shows good result on the BioId database. Camadelli attribute this behavior to the major similarity of BioId database to the images used to train CLM. Since Campadelli's method show better results, therefore I would like to implement the method. The method is described in detail in the next section.

5 Precise Eye localization trough a general-to-specific model definition [In Detail] -

Here I will discuss this approach I detail because in my next semester I will be implementing this algorithm.

5.1 Pattern definition and feature selection -

According to the author wavelet pattern for the eye pattern is more favorable the then direct representation as it leads to a smaller generalization error. Haar-like wavelets permit to describe the pattern in terms of luminance changes at different frequencies, at different positions and along different orientations.

Before the wavelet decomposition, each image undergoes an illumination normalization process and then it is reduced to 16 x 16 pixels. This is a trade off between the necessities to maintain the low computational cost and to have sufficient details to learn the pattern shape. The decomposition is realized via an overcomplete bi-dimensional FWT (Fast Wavelet Transform). A wavelet coefficient $d^o_{j,k1,k2}$ is identified by four parameters: j is called the details level and relates to the size of the window over which the coefficient is calculated (hence it regulates the frequency); $(k1, k2)$ is called the shift and relates to the position of the coefficient within the image; $o \epsilon \{horizontal, vertical, diagonal\}$ determines the orientation of the edge that is tested for presence. The set B_j of all $d^o_{j,k1,k2}$ of a certain level j is called the band of level j.

Mean illumination of the image is discarded producing a sort of illumination normalization of the pattern examples. Clearly band corresponding to the highest frequency is crucial for the precision of the eye localizer so it can be dropped in the case of eye detector at an expense of coarser localization. By doing this author manage to both to specialize the two classifier by specifying a different maximum detail level. Real selection process is done by the normalization step: Take a set L of eye pattern images and decompose each $l \epsilon L$ in its wavelet coefficients $d^o_{j,k1,k2}(l)$. For each coefficient calculate its mean value -

$$\overline{d^o}_{j,k_1,k_2} = \frac{\sum_{l=1}^{|L|} |d^o_{j,k_1,k_2}(l)|}{|L|}$$

And normalize it with respect to the average mean of its band; then -

$$\tilde{d^o}_{j,k_1,k_2} = \frac{\overline{d^o}_{j,k_1,k_2}}{m_j}, \quad \text{where } m_j = \frac{\sum_{k_1}\sum_{k_2}\sum_o \overline{d^o}_{j,k_1,k_2}}{|B_j|},$$

The normalized coefficients $d^o_{j,k1,k2}(l)$ ¿ 0 can be interpreted as follows -

$$\tilde{d^o}_{j,k_1,k_2} \begin{cases} \sim 1 & \Rightarrow \text{no regularity} \\ \ll 1 & \Rightarrow \text{systematic uniformity} \quad (C^-) \\ \gg 1 & \Rightarrow \text{systematic variation} \quad (C^+) \end{cases}$$

Hence the normalization distinguishes the two sub-categories of coefficients that can be ordered separately. C+ class gathers the coefficients that capture the edge structure of the pattern; C- class contains the coefficients that represent the significant absence of the edges.

Once ordered the normalized coefficients an error function can be defined to drive the selection process.

$$w = \arg \min_{\substack{w = w^+ \cup w^-, \\ w^+ \subseteq C^+,\ w^- \subseteq C^-}} ||E - E_w||^2 + \alpha \cdot ||E_w - U||^2$$

Here E is the mean eye pattern, U is a uniform pattern (with all pixels set to the mean luminance of E) and E_w is the reconstruction obtained by retaining the set w of the wavelet. First term of the objective function represents the error made by the construction, while the second term intends to bound amount of details that are being added to the pattern representation. Set $w = w^+ \, U \, w^-$ is selected such that it corresponds to a local minimum of the objective function, and the ratio $|w^+|/|C^+|$ roughly equals $|w^-|/|C^-|$.

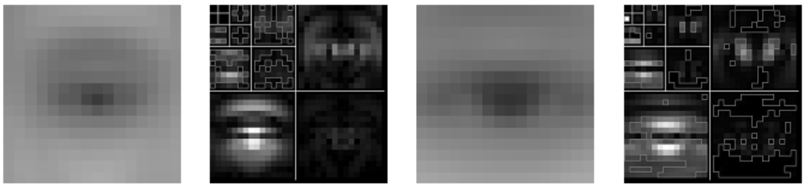

Figure 1: From left to right: the mean eye pattern, its wavelet decomposition and the selected features (red contour) for the SVM-1 and the SVM-2 respectively. High intensities correspond to strong edges, low intensities indicate uniform regions.

For the eye detector it retain less than 100 coefficients, while for the eye localizer it keep about 300 coefficients, therefore the application of the second classifier is mush costly than the first one.

The differentiation of the two SVM is also achieved by choosing a suitable sample of training examples for each of them: SVM1 must distinguish the global eye shape from other features, especially those found inside the face. The positive class is built to contain the eye images cropped to a size equal to the inter-ocular distance. The negative class is built by the other facial features and by some examples extracted from the background. The SVM2 is presented with positive examples that correspond to a smaller receptive field (half of the eye pattern previously defined) and with negative examples that are generated by small, random displacement of the sub-images used for the extraction of the positive examples.

5.2 Localization Technique -

Although the position and scale of the face in the image are received as input to the algorithm, we know that any automatic face detector is subject to a certain error distribution on the size of the outputs, besides the presence of false positives. The first uncertainty is accounted by considering a range of three scales at which to search eyes, and eye detector is used as a validator of the face to discard false positives.

Given a region output by the face detector, the evaluation of a candidate point within that region comes to evaluating three examples centered in it: the one at the inferred scale (X_P), plus two examples extracted according to small underestimation and a small overestimation of that scale (X_P^- and X_P^+). The size X_P^- and X_P^+ are chosen to account for an estimation of the face size that is between half and twice the true size. If SVM$(x) = 0$ is the equation of the hyperplane that separates the two classes of positive and negative examples, then SVM(x) can be treated as a measure of the confidence with which SVM classifies the example x. Thus the function can be described as the strength of the candidate P -

$$\rho(P) = SVM_1(\mathbf{x}_P) + SVM_1(\mathbf{x}_P^-) + SVM_1(\mathbf{x}_P^+)$$

Authors proceed by evaluating $\rho(P)$ over a small subset of points in the face region: first they identify the points that lie on the edges, then they subsample them that depends on the scale of the face region; they consider as point candidates the ones for which $\rho(P)$ is greater than 0, and then group them according to their proximity in the image; each group of point candidates is then represented by its centroid (the eye candidate) obtained weighting each point P with its $\rho(P)$. This last step strengthens the eye detection, making it more stable.

At last, they refine the results of the detection by applying the SVM-2 within a small neighborhood of the found positions. The scale considered by the SVM-2 should approximate half of inter-ocular distance and is obtained as follows -

$$\frac{1}{2} \times \frac{\sum_{\mathbf{x} \in \{\mathbf{x}_P, \mathbf{x}_P^+, \mathbf{x}_P^-\}} [\Theta(SVM_1(\mathbf{x})) \times scale(\mathbf{x})]}{3} \quad \text{where} \quad \Theta(z) = \begin{cases} z & \text{if } z > 0 \\ 0 & \text{if } z \leq 0 \end{cases}$$

The search for the candidates that gives the highest response according to the SVM-2 proceeds analogously as in the eye detection module.

6 Work done this semester -

- Reading papers for better understanding the problem
- Deciding the Algorithm for implementation
- Downloading the face databases

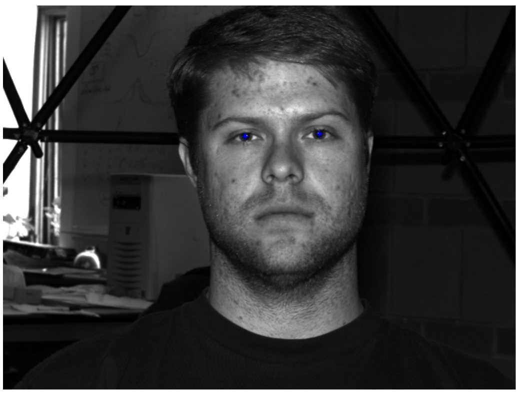

Figure 1: Face image form Yale database [5], eye positions are given in text files

Figure 2:Face imags from BioId [7] and Feret [6] database respectively, Eye positions are given in text files

7 Future Work

Implementation of the above algorithm -

- Implementation of the feature selection module using Haar wavelets decomposition on the output of Viola-Jones face detector
- Implementation of the eye localizer and eye detector using support vector machine
- Training of the module
- Eye detection accuracy
- Comparison with other Algorithms

References

[1] Campadelli, P., Lanzarotti, R., and Lipori, G. Eye localization: a survey. *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue* (2007), 234.

[2] Campadelli, P., Lanzarotti, R., Lipori, G., and di Milano, U. Precise eye localization through a general-to-specific model definition. In *Proceedings of the British Machine Vision Conference* (2006), vol. 1, pp. 187–196.

[3] Cristinacce, D., and Cootes, T. Feature detection and tracking with constrained local models. In *Proceedings of the 17th British Machine Vision Conference, Edinburgh, UK* (2006), pp. 929–938.

[4] Everingham, M., and Zisserman, A. Regression and classification approaches to eye localization in face images. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition* (2006), pp. 441–448.

[5] http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html. The yale database.

[6] http://face.nist.gov/colorferet/request.html. The feret database.

[7] http://www.bioid.com/support/downloads/software/bioid-face database.html. The bioid database.

[8] Riopka, T., and Boult, T. The eyes have it. In *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications* (2003), pp. 9–16.

[9] Tang, X., Ou, Z., Su, T., Sun, H., and Zhao, P. Robust precise eye location by adaboost and svm techniques. In *Proceedings of the International Symposium on Neural Networks* (2005), pp. 93–98.

[10] Wang, P., Green, M., Ji, Q., and Wayman, J. Automatic eye detection and its validation. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops* (2005), p. 164.