BTP Report

Eye Localization using Haar Wavelets and Cascaded Support Vector Machine

Submitted by:
Vishal Agrawal
Y6541 (agrawal@iitk.ac.in)
Guide: Prof. Amitabha Mukherjee/ Dr. Simant Dubey
Final Year Under-Graduate
Dept. of Computer Science and Engineering
IIT Kanpur – 208016
INDIA

Table of Contents

Abstract:

In the last two decades large amount of work has been done in the field of the face recognition, but the interest in this field is still high. Nowadays interest is in the field of making the face recognition completely automatic, which requires automatic localization of the facial features (eyes, nose, mouth, chin etc). Many recent papers[6,7] deal with the uncontrolled environment conditions, but they are not fully automatic, in some way they rely on the manual annotations of the facial features, thus limiting its use in the real world. Therefore for real world face recognition application we need a facial features localization technique.

In this report I will present a fully automated Eye localization technique in 2D images using properly selected Haar Wavelet coefficients and the two cascaded SVMs (Support Vector Machine).

Motivation:

Symmetry of the eyes makes them one of the most important facial features out of several other facial features [2, 3, 4]. Once the eye positions are detected then the other facial features can be detected very easily. Moreover accuracy of the face recognition system depends on the accurate localization of the facial features, these facial features either are used a classier directly or for normalizing the images. Eye separation generally does not change significantly with facial expression, nor with up and down movement of the face, therefore separation of the eyes, is generally used for normalizing the face images. Face alignment is also determined using the positions of eye. This makes eyes feature an extremely useful in face recognition application.

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

Eye Localization Approach:

This approach [1] use two SVMs, trained on the properly selected Over-complete Haar wavelet coefficients. Over-complete Haar wavelet coefficients permit to describe the pattern in terms of luminance changes at different level that corresponds to frequency, at different positions and along different orientations.

In the following sections I present the technical details of the approach.

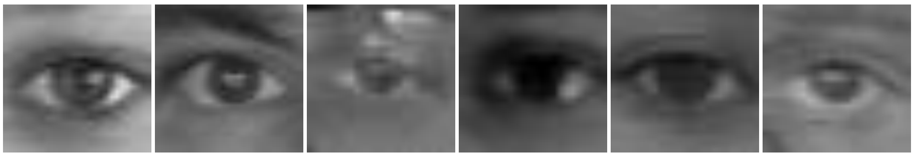
Training set:

First SVM (SVM1) is meant for distinguishing the eye shape from the other features, especially those found inside the face. Thus SVM1 works as an eye detector. Therefore the positive class contains the eye images cropped from the face at the size of inter-ocular distance, centered at the center of eyes. Negative images contain the other facial features, and some background images. (8 – 11 negative images for each positive.)

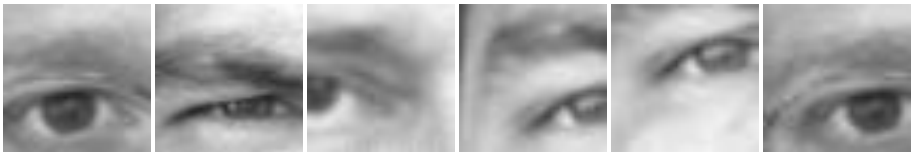
Second SVM (SVM2) works as an eye localizer, it works in cascade, and its task is to calculate the precise position of the eyes center. Positive class is made of the eye images as in the SVM1, but this time the size of the images is half of the size of the images in the SVM1. Negative class is generated by the small random displacement of the sub-images used for the extraction of the positive examples. (10 negative images for each positive.)

SVM1 Positive Images

SVM1 Negative Images



SVM2 Positive Images



SVM2 Negative Images

Feature selection:

Before applying the wavelet decomposition on an image, image goes to an illumination normalization process. Then the size of the image is reduced to 16x16. 16x16 is a trade-off to maintain the low computational cost and to have the sufficient details to learn the pattern shape. The decomposition is achieved using the Over-complete Haar wavelet transform. A wavelet coefficient $d^{o}_{j,k1,k2}$ is identified by the four parameters: j is the detail level, hence it corresponds to the frequency, $(k1, k2)$ relates to the position of the coefficients within the image, o belongs to the $\{horizontal, vertical, diagonal\}$. The set B_j of all $d^{o}_{j,k1,k2}$ of a certain level is the called the detail level j.

Over-complete wavelet decomposition gives us total 852 coefficients for each image.

$|B_o| = 3 * (15 * 15) = 675$
$|B_1| = 3 * (7 * 7) = 147$
$|B_2| = 3 * (3 * 3) = 27$

$|B_3| = 3 * (1 * 1) = 3$
$Total = |B_o| + |B_1| + |B_2| + |B_3| = 852$

In order to select the features from these 852 coefficients, we need a way to calculate the relative importance of the coefficients. This is done by the normalization step followed by an error function.

Normalization step

Let say L is the set of eye images (positive images) in the training set, decompose each $l \in L$ in its wavelet coefficients $d^o_{j,k1,k2}(l)$. For each $d^o_{j,k1,k2}$ calculate its mean value in the set L –

$$\bar{d}^o_{j,k1,k2} = \frac{\sum_{l \in L} |d^o_{j,k1,k2}(l)|}{|L|}$$

Now normalize it with respect to the average mean of its band –

$$\tilde{d}^o_{j,k1,k2} = \frac{\bar{d}^o_{j,k1,k2}}{m_j} \quad where \ \ m_j = \frac{\sum_{k1} \sum_{k2} \sum_o \bar{d}^o_{j,k1,k2}}{|B_j|}$$

$\tilde{d}^o_{j,k1,k2}$ represents the normalized coefficients that can now be order to access the relative importance of the coefficients.

Now,

$$E\left[\sum_{k1} \sum_{k2} \sum_o \tilde{d}^o_{j,k1,k2}\right] = \sum_{k1} \sum_{k2} \sum_o E\left[\frac{\bar{d}^o_{j,k1,k2}}{m_j}\right]$$

$$E\left[\sum_{k1} \sum_{k2} \sum_o \tilde{d}^o_{j,k1,k2}\right] \approx \frac{\sum_{k1} \sum_{k2} \sum_o E\left[\bar{d}^o_{j,k1,k2}\right] * |B_j|}{E\left[\sum_{k1} \sum_{k2} \sum_o \bar{d}^o_{j,k1,k2}\right]} = |B_j|$$

Since the expected value of the sum of $\tilde{d}^o_{j,k1,k2}$ in a particular band is approximately equal to the cardinality of the band, the normalized coefficients $\tilde{d}^o_{j,k1,k2} > 0$ can be interpreted as follows-

$$\tilde{d}^o_{j,k1,k2} \quad \begin{cases} \sim 1 & => no\ regularity \\ \ll 1 & => systematic\ uniformity \quad (C^-) \\ \gg 1 & => systematic\ variation \quad\ \ (C^+) \end{cases}$$

Hence the normalization allows us to distinguish the normalized coefficients in two classes: C^- and C^+. C^- corresponds to the absence of edges and C^+corresponds to the edge structure of the pattern.

Error function

After obtaining the ordered coefficients, we define an error function to drive the selection of features.

$$\mathrm{w} = \min_{\substack{\mathrm{w}=\mathrm{w}^{+}\cup\mathrm{w}^{-}, \\ \mathrm{w}^{+}\subseteq\mathrm{C}^{+},\mathrm{w}^{-}\subseteq\mathrm{C}^{-}}} ||\mathrm{E}-\mathrm{E}_{\mathrm{w}}||^{2} + \alpha * ||\mathrm{E}_{\mathrm{w}} - \mathrm{U}||^{2}$$

Where E is the mean eye pattern. E_w is the reconstructed pattern using the set w of the coefficients, U is the uniform pattern. $||\mathrm{E}-\mathrm{E}_{\mathrm{W}}||^2$ is the measure of the error made by the reconstructed pattern, while $||\mathrm{E}-\mathrm{E}_{\mathrm{W}}||^2$ is the measure of information that we have added in the reconstructed image. α is a trade-off to balance these two opposite goals.

Selection of set w should be such that $\frac{|w^{+}|}{|C^{+}|}$ roughly equals $\frac{|w^{-}|}{|C^{-}|}$

Training

For the SVM1 we discard the features that are in the highest level wavelet coefficients, because highest level wavelet coefficients are important for the precise localization of the eye and the SVM1 must be general and robust even if at the expense of coarser localization. While in the case of SVM2 we retain all the features.

For each image in the training set (I have used 2380 positives for training), we write the selected wavelet coefficients into a file. This is our actual training data.

I have used the libsvm library for the SVM.

Localization Technique

First we apply an automatic face detector to get the face region from the test image (I have used the OpenCV's version of the Viola-Jones face detector). Face detector gives us the face region and the scale of the face. But, any face detector is subject to a certain error distribution on the size of its outputs, besides the presence of the possible false negatives. SVM1 also works as a face validator. To deal with the first uncertainty in SVM1 we search for the eyes in the face at three different scales. Evaluation of a point P in the face region comes to evaluating three examples centered in it: one at the actual inferred scale(X_P) and other two at the small underestimation (X_P^-) and the small overestimation (X_P^+) of the scale.

Let say $SVM1(X) = 0$ is equation of the hyper-plane by which SVM1 separates the eyes and non-eyes. Thus we calculate the strength of the point P by which it is classified as an eye as follows –

$$\rho(P) = SVM1(X_P) + SVM1(X_P^-) + SVM1(X_P^+)$$

We proceed by evaluating the strength over small subsets of the points in the face region. We consider those points as a potential eye center for which we have $\rho(P) > 0$.
Then we group the points based on their proximity in the image, each group is then represented by the centroid of the group, which is calculated by the weighted sum of the point P with its $\rho(P)$.

In the ideal case we should have only two points, if we have more than two points than it can be handled by using the verticality of the case as follows –

$$\{C_a, C_b\} = \max \frac{SVM1(C_i) * SVM1(C_j)}{|C_{iy} - C_{jy}|}$$

Thus at the end of the SVM1 we have only two positions of the eyes center. At last we refine the results by using SVM2 within a small neighbourhood of the detected positions. For the SVM2 we calculate the scale as follows -

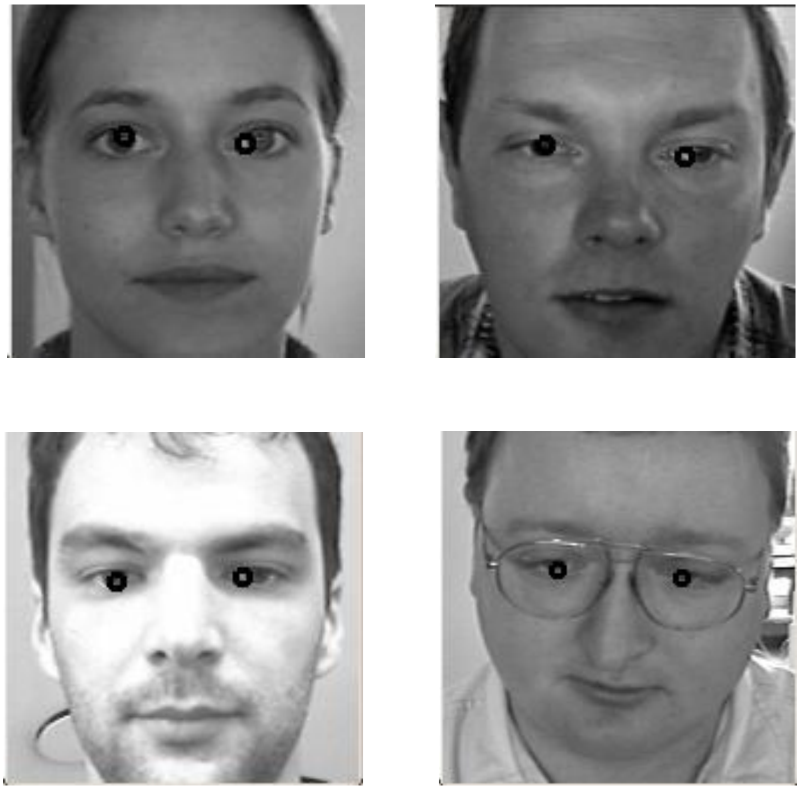
$$\frac{1}{2} * \frac{\sum_{X \in X_p, X_P^-, X_P^+} \theta\big(SVM1(X)\big) * scale(X)}{3}$$

Where

$$\theta(z) = \begin{cases} z & if\ z > 0 \\ 0 & otherwise \end{cases}$$

Results:

Images showed below shows the detected eyes positions in the face region by the small black circle.



Detected Eye Positions

Conclusion:

The method follows the top-down approach. Starting from the face detection, it applies the SVM1 which gives a coarser location of the eyes center. Then it applies the SVM2 to improve the localization.
In the recent years many research works have pointed that the precise and automatic eye localization is a crucial step for the real world biometric applications.

References:

1. 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.

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

3. Riopka T., Boult T., The eyes have it: *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications* (2003), pp. 9-16

4. G. Givens, J.R. Beveridge, B.A. Draper, P. Grother, and P.J. Phillips. *How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. Proceedings of IEEE on Computer Vision and Pattern Recognition*, 2004.

5. 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), pp. 164

6. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkaba. *Discriminative common vectors for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1): 4–13, 2005.

7. K.I. Chang, K.W. Bowyer, and P.J. Flynn. *An evaluation of multimodal 2d+3d face biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005

8. Yale Database. *http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html*

9. Feret Database. http://face.nist.gov/colorferet/request.html

10. BioId Database http://www.bioid.com/support/downloads/software/bioid-face-database.html