Multi-stage Combination of Geometric and Colorimetric Detectors for Eyes Localization

Maurice Milgram, Rachid Belaroussi, and Lionel Prevost

LISIF Université Pierre et Marie Curie BC 252
4, Place Jussieu, 75252 Paris Cedex 05, France
maurice.milgram@upmc.fr,
rachid.belaroussi@gmail.com, lionel.prevost@upmc.fr

Abstract. We present in this paper a method for the localization of the eyes in a facial image. This method works on color images, applying the so called Chinese Transformation (CT) on edge pixels to detect local symmetry. The CT is combined with a skin color model based on a modified Gaussian Mixture Model (GMM). The CT and the modified GMM give us a small rectangular area containing one eye with a very high probability. This rectangle is then processed to find the precise position of the eye, using four sources of information: a darkness measure, a circle finder, a “not skin” finder and a position information. Experimental results on a large database are presented on nearly 1000 faces from the ECU database.

1 Introduction

We present in this paper an algorithm for locating the eye pair in a facial image. Face detection/localization has become a very important issue for applications such as videoconferencing, biometrics, human detection, digital picture retrieval, etc. This issue is connected to several sub-problems: detection of facial landmarks (eyes, mouth, etc..), gaze detection and tracking, pose estimation, facial expressions analysis.

Detecting faces often starts with the detection of face appearance (pattern) or color, normalizing the image to take into account illumination and geometrical variations. Facial landmarks bring useful information to cope with the huge variability of face appearance. In this spirit, eye detection is a very important step in face detection. It can help to confirm or reject a hypothesis, decreasing the false detection rate.

We find two generic approaches in eye detection:

- Binary decision where the question is: "is this precise scanning window an eye or not?". In this approach, the problem of detection is broken into a large set of decision problems.
- Localization where ones try to locate eyes in a face image (or sub-image), supposing that this image contains at least one visible eye. In this more holistic approach, we can use global information.

Our method belongs to the second family but can also make use of some local properties.

Huang and Wechsler [8] used wavelets to represent eyes and RBF networks to perform the classification of facial regions as eye / non-eye. Gabor wavelets are often used to process face images and iris coding [5]. In [3] on the detection of “genuine smiles”, authors indicate that Gabor filters representations of images are highly effective for face recognition and expression analysis. Some authors [7] propose to extract feature points from faces with help of edge and corner points. They use classical image analysis techniques (binarization, projection, gray level morphology) to find these feature points in a pure bottom up way. A method to locate eyes in human face images using fractal dimensions is presented in [10]. Sometimes, authors try to find a specific feature inside the eye: [2] introduce a method to detect the white visible portion of the eyeball, the sclera which color is different from the surrounding area.

The paper is organized as follows. We present the database and the pre-processing in section 2. In section 3, we introduce our Chinese Transform. Next section is devoted to a probabilistic method for improving our detection and allowing an extrapolation of the second eye if only one has been detected. Section 5 describes a refinement technique for the y coordinate. Experimental results are presented in section 6. The last section is devoted to our concluding discussion.

2 Database, Pre-processing and Normalization

ECU database [11] contains about 3000 pictures with one face or more. Pictures are of various complexity, resolution and illumination. We used the first half of the database (training set) for our statistic estimations and parameters tuning. The whole database was used to get experimental results and no significant bias was observed between the two halves of the database. Face image, after the location of the head boundary, is transform in a grayscale image by simply averaging the three RGB values without taking into account of prior knowledge on the skin color like in [9]. This makes the method robust with respect to illumination variations and allows the treatment of monochrome pictures (for example close infra-red). We assume in our paper that the locating of the head boundary is done to focus on eye location. In order to normalize the images from the radiometric point of view, we chose a linear stretching of the gray-levels dynamic. This transformation gave better results than the well-known histogram equalization.

To find the orientation of edge pixel gradient, a circular mask is used for low pass filtering. Then, we classically determine the 2 components (G_x and G_y) of the gradient by using a simple Roberts operator. We are interested only in the direction of the gradient ($\theta = \arctan(G_y/G_x)$). The gradient magnitude is only used to select edge pixels. It should be recalled that the gray level dynamic is always between 0 and 255 because of the initial stretching. The orientation θ is quantified on N values ($N=8$ for us). We will call *ORIENX* the map of the quantified orientations thus obtained; this map can be divided into $N+1$ sets corresponding to orientations $0, 1, \dots, N-1$ and one more set corresponding to non-edge pixels.

3 First Stage: Coarse Eye Localization

3.1 Chinese Transform

CT is a global transformation which input is the gradient orientations of the edge pixels of the image. It provides as output two feature maps (or cumulative arrays). The main feature map (so called *VOTE*) gives information on the centers of the eyes. The basic idea is that the eyes occupy, including the eyebrows, a zone which is darker than its environment and has an elliptic shape. Because of its property of approximate central symmetry, the eye center is the middle of many segments joining 2 opposite points of the boundary and these pairs of points having opposite gradient orientations (Fig 1. shows an example). The CT is quite different but related to the work presented in [12] on generalized symmetry. Unlike the Hough Transform, the feature space of the CT is simply the image. Although CT looks like the GHT (Generalized Hough Transform which uses the gradient), it strongly differs from it because one CT can detect at the same time ellipses of any size and eccentricity. It is based on the central symmetry property of ellipses.

Fig. 1. Applying CT to an ellipse. Left: gradient orientation of edges. Right: resulting *VOTE*.

To start, the general principle of CT consists in determining, for each pair of pixels (M, M') with orientations $O = \text{ORIENX}(M), O' = \text{ORIENX}(M')$ ranging within $[0 ; N-1]$ if they are part of the voter set *VS* or not. In our case, the criterion of membership is based on the relative position of M and M' and on the condition $O = 2$ and $O' = 6$ (in other words $O = \text{North}$ and $O' = \text{South}$). The relative position of the points $M(x, y)$ and $M'(x', y')$ must check the following geometrical constraints:

- (1) $y - y' \geq \delta_{\text{VERT}}$
- (2) $45^\circ < \alpha < 135^\circ$ where $\alpha = \arctan \frac{y - y'}{x - x'}$
- (3) $\|MM'\| < \delta_{\text{DIST}}$

These 3 conditions indicate that (1) M is above M' , (2) the vector MM' is not horizontal and (3) points M and M' are not too distant. One can notice that the first condition does not involve the second one since the points can be relatively distant on the horizontal axis. The values $\delta_{\text{VERT}} = 0.08H$ and $\delta_{\text{DIST}} = 0.25H$ are taken from experimental results (the parameter H is the image height). We define and initialize to zero two arrays *VOTE* and *DIAM* which have the size of *ORIENX* (orientation map). For each pair (M, M') of elements of the voter set *VS*, we increment the value of *VOTE*(P) where P is the middle of the segment $[M, M']$. The CT name comes from

the fact that China is sometimes called the "empire of the middle". To preserve an information about the distance between the 2 points voters for a point P , we store in $DIAM(P)$ (for diameter) the average of the distance $d(M, M')$ where P is the middle of $[M, M']$ (the CT name comes from the fact that China is sometimes called the "middle empire"). This information will be also used to determine the size of a rectangular window around each eye. The result provided by the CT is thus a pair of maps: *VOTE* and *DIAM*. True eyes locations correspond fairly well to the pixels with maximum number of votes. To eliminate most of parasitic local maxima, we carry out a smoothing of the *VOTE* map. Nevertheless, we need additional cues to find accurately the eyes location.

3.2 Modified GMM Skin Color Model

The skin color is often used as a cue for detecting, localizing and tracking faces. Several classes of skin color models have been proposed: non parametric, parametric and neural. GMM belongs to the category of parametric estimators and are very popular, especially for the skin/no-skin discrimination. In [4,6], authors discuss the choice for the best chromatic space (RGB, HS, YcbCr, YIQ, CIE XYZ, ab, etc.), the training algorithm (most of them use the Expectation-Maximization algorithm) used to learn parameters (weights, means and covariance matrix) and to determine the complexity (number of components).

We decided to learn a skin color model for each face, hence in a non-supervised approach. The idea is that in general, eyes are in the upper part of the face. We learn our parametric model on the lower part of the face and then apply it on the whole face, looking for pixels that are not well accepted by the model. EM algorithm has proved to be efficient when using a large amount of data for training. In our case, we prefer a simpler method because we have to deal with a small number of skin-pixels. This number is small because we work on one single image and also because we sub-sample (1 out of 9) the image to eliminate (or to decrease the weight of) irrelevant regions (mouth, nostrils, beard). So we use a simple K-means algorithm with Euclidean distance to get the best center of each Gaussian. With each center, we have a specific subset of training examples that are closer to this center than to the other. We estimate Gaussian parameters (means and covariance matrix) for each subset.

Several experiments lead us to choose the RGB colorimetric space that gives better results than the YCbCr one. The number of Gaussian was set to 3 and we perform 5 iterations of K-means with an initialization based on a uniform spatial repartition of pixels (after sub-sampling and working only on the lower part of the face image). Applying this model to the whole face image produces the *Skin_map* containing the Mahalanobis distance of each pixel (in the RGB space) to the best center found with the K-means algorithm.

3.3 Combination

We have now 2 maps: the Smoothed *VOTE* accumulator of the Chinese Transform and the *Skin_map*. We search first all local minimum of *Skin_map*. Let LM be this set. We believe that eyes locations are probably close to one of these minima. We simply take the point (AM) with the absolute maximum of the smoothed vote accumulator over LM pixels. We have found experimentally that a rectangle centered on the point

AM and with a width and height of $0.25ImSize$ gives a probability of 99%. $ImSize$ is the mean of the width and the height of the image computed over the whole training set. To find the rectangle for the second eye, we set to zero all the votes of the first rectangle in the vote accumulator; then we search again the absolute maximum of this modified array.

Fig. 2. Combination of CT with the modified GMM

4 Second Stage: Fine Eye Localization

We have found a small rectangle for each eye and we have now to refine this area to locate more precisely each eye. Computing 4 maps will do this:

4.1 Darkness Measure

Since we have already found a small area containing the eyes, we can exploit the fact that pupils are dark regions. Empirical tests showed that a threshold of 25% is good. We compute a darkness map with the formula (4), for each pixel (i,j) with gray level $g(i,j)$:

$$(4) \quad Dark_map(i,j) = U(\delta - g(i,j)) + \frac{10U(g(i,j) - \delta)}{g(i,j) - \delta + 10}$$

where $U(x)$ is the usual threshold function, δ is the value to get 25% of pixel gray levels below. $Dark_map$ takes the value 1 for all pixels below δ ; for other pixels, it takes a decreasing value as the gray level increases. With $\delta=55$, $Dark_map=0.05$ for white pixels (gray level=255), and $Dark_map=0.5$ for the gray level 65. So it is a fast decreasing function from the value 1 as the gray level increases from δ value.

4.2 Pupil Finder

We compute the classical Generalized Hough Transform for Circle (GHTC) using the intensity gradient orientation. Each edge pixel votes for the circle center in the opposite direction of the gradient, at a distance equal to the (hypothetical) radius. This GHTC is applied for each possible value of the radius. We find that the mean radius of the pupil is $0.14L$ where L is the width of the rectangle. So we take $R = 0.14L$ and the set of possible radii is the interval $[0.7R \ 1.3R]$. After the voting process, values are normalized in the range $[0 \ 1]$ and give us the *Hough_map*.

4.3 Local Skin Color Model

The global skin model computed in the previous section is no more accurate for the rectangle. We compute again a local skin color model with sub-sampling (factor is set to $\frac{1}{2}$) and with only one Gaussian. We obtain the *Skin_map*.

4.4 Position Information

Rectangle provided by the first stage of our method gives a rather centered position for the “true” position of the pupil. It means that there is more information provided by this step than a simple rectangle, implicitly representing a uniform density probability. We build the *Position_map* as 2D triangular function with value 1 in the center of the rectangle and 0 just on the boundary of the rectangle.

The 4 maps are simply multiplied to get a *Fusion_map*. It means that we consider these four maps equivalent and we make a kind of “fuzzy logical and” of these four information. Other kinds of fusions were tested but none outperformed the simple multiplication. The position of the center of the eye (see discussion below about the ground truth) is determinate as the location of the absolute maximum of the *Fusion_map* (figure 3) within each sub-window.

Fig. 3. Hough map (a), skin map (b), dark map (c), fusion map (d) and localization (e) (red: ground truth, green: localization)

5 Results and Discussion

Recall that we have taken all faces of the ECU database, even with closed eyes, glasses, hairs on face, unusual pose. We just select faces with more than 150 pixels width and height (32% of the faces are selected with a mean width of 218 pixels). The ground truth was done manually: operators click on the pupil’s center of each eye, even if person are wearing glasses, even with reflections. We have checked this ground truth and it appears that there is a small dispersion among operator labelling. Moreover, our method is not really a “pupil finder” but rather a “eye finder”. One can see that, sometimes, pupils are far from the eye center or even are not visible (closed eye). If the distance is smaller than 10 pixels (true with probability 0.9), it means that the ground truth is within a square of 14×14 pixels centered on the detection. This

value of 14 pixels is about 6.4 % of the average width of the selected faces. For these 923 faces, we compute the error i.e the distance (in pixels) between the ground truth and the detection provided by our algorithm. The histogram is presented (figure 4). We obtain several statistics for the error (table 1). We can summarize our results like this: for 90% of the faces, we accurately find the eyes and in 50%, we find exactly the pupil center. We present in figure 5 some examples of eye localization.

Fig. 4. Error (in pixel) histogram for 923 faces

Table 1. Localization error

Median	Mean	P(Error<10)	P(Error<15)	P(Error<20)
3.6056	5.0609	0.9147	0.9579	0.9761

We have done some experiments to evaluate the processing time. These measure have been done on 20 images with a MatLab 6.5 environment. The average computing time is 9.1 sec per image for the total detection, processor is a Pentium 4 at 2.6 Ghz.

6 Conclusion and Prospects

We have presented a system that detects and localizes precisely eyes in a facial color image. It is a two-stage combination of several detectors based on the colorimetric properties of face and eye and the geometric properties of eye and pupil. We introduce a new cumulative transform that can detect central symmetry. Experimental results show the accuracy of our method.

As we developed in our lab an automatic face localizer [1], eye detector should be integrated as a part of the whole system to help, confirm or reject a hypothesis in order to decrease the false detection rate.

Fig. 5. Examples of localization (red: ground truth, green: localization)

References

- [1] Belaroussi, R. , Prevost, L. & Milgram, M.: Combining model-based classifiers for face localization, to appear in Proc. of IAPR Conf. on Machine Vision & Application (2005)
- [2] Betke, M., Mullally, W. & Magee, J.: Active detection of eye sclera, IEEE CVPR Workshop on Human Modeling, Analysis and Synthesis (2000)
- [3] Braathen, B. and Bartlett, M. S. Littlewort-Ford G. & Movellan J. R.: 3-D head pose estimation from video by nonlinear stochastic particle filtering, UCSD MPLab TR 2001.05 (2001)
- [4] Caetano, T.S. & Barone, D.A.C.: A probabilistic model for the human skin color, IEEE Int. Conf. on Image Analysis and Processing, Int. Conf. On Image Analysis & Processing (2001) 279-283
- [5] Daugman, J.G.: Complete Discrete 2-D Gabor Transform by Neural Networks for Image analysis and Compression, IEEE Trans. ASSP, Vol. 36, n°7 (1988) 1169-1179
- [6] Garcia, C. & Tziritas, T.: Face detection using quantized skin color regions merging and wavelet packet analysis, IEEE Trans. on Multimedia, Vol. 3, (1999) 264-277
- [7] Gu, H., Su, G. and Du, C.: Feature points extraction from face, Image and Vision Computing New Zealand (2003)
- [8] Huang, J. & Wechsler, H.: Eye location using genetic algorithm, Int. Conf. On Audio and Video-Based Person Authentication (1999)
- [9] Hsu, R.L., Abdel-Mottaleb, M. & Jain, A.K.: Face detection in color images, IEEE Trans. PAMI, Vol. 24, n°5 (2002) 696-706
- [10] Lin, K.H., Lam, K.M. & Siu, W.-C.: Locating the eye in human face images using fractal dimensions, IEEE Proc.Vis.Image Signal Process., Vol. 148, n°6 (2001)
- [11] Phung, S.L., Bouzerdoun, A. & Chai, D., Skin segmentation using color pixel classification: Analysis and comparison, IEEE Trans. On PAMI, to be published in 2005
- [12] Reisfeld, D. & Yeshuran, T., Robust detection of facial features by generalized symmetry, IEEE Int. Conf. on Pattern Recognition (1992) 117-102