FACIAL FEATURE EXTRACTION USING ADAPTIVE HOUGH TRANSFORM, TEMPLATE MATCHING AND ACTIVE CONTOUR MODELS

Athanasios Nikolaidis *Constantine Kotropoulos* *Ioannis Pitas*

Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki 540 06, Greece.
E-mail: {nikola, costas, pitas}@zeus.csd.auth.gr

Abstract: The present paper describes an extention to the methods proposed in [1] for the extraction of facial features with the ultimate goal to be used in defining a sufficient set of distances between them so that a unique description of the structure of a face is obtained. Eyebrows, eyes, nostrils, mouth, cheeks and chin are considered as interesting features. Candidates for eyes, nostrils and mouth are determined by searching for minima and maxima in the $x-$ and $y-$ projections of the greylevel relief. Candidates for cheeks and chin are determined by performing adaptive hough transform on a relevant subimage defined according to the position of the eyes and mouth and the ellipse containing the main connected component of the image. A deforming technique is also applied to the ellipse representing the main face region, in order to acquire a more accurate model of the face. Candidates for eyebrows are determined by adapting a proper greylevel template to an area restricted by the position of the eyes. The algorithms presented were tested on a set of 37 different color images, containing features as beard, glasses and changing facial expressions.

1. INTRODUCTION

One of the key problems in building automated systems that perform face recognition tasks is face detection and facial feature extraction. Many algorithms have been proposed for face detection in still images that are based on texture, depth, shape and colour information of them. For example, face localization can be based on the observation that human faces are characterized by their oval shape and skin-colour [1].

Another very attractive approach for face detection is based on multiresolution images (also known as mosaic images) attempting to detect a facial region at a coarse resolution and, subsequently, to validate the outcome by detecting facial features at the next resolution [2], [3]. A critical survey on face recognition can be found in [4].

The major difficulties encountered in face recognition are due to variations in luminance, facial expressions, visual angles and other features such as glasses, beard, etc. This leads to a need for imposing several constraints on the set of images on which the specific algorithm is tested.

Several techniques for facial feature extraction have been proposed in the late years. They can be divided into two broad categories: template-based techniques and feature-based ones. An interesting comparison between the two categories can be found in [5].

Experimental research through years proved that a fusion of the two basic techniques usually provides more stable results than using each technique alone. The success of a specific technique depends on the nature of the feature to be extracted. In the present paper, the features under consideration are the eyebrows, eyes, nostrils, mouth, cheeks and chin. The eyes, the nostrils and the mouth are determined by searching for minima and maxima in the $x-$ and $y-$ projections of the greylevel relief (see [1],[6]). In this paper, the following methods are proposed: i) the extraction of cheeks and chin by performing an Adaptive Hough Transform (AHT) on a relevant subimage defined according to the ellipse containing the main connected component of the image, ii) the extraction of upper eyebrow edges using a template-based technique that adapts a proper greylevel mask to an area restricted by the position of the eyes, and iii) the accurate extraction of the face contour by means of dynamic deformation of active contours (e.g. snakes), compensating for the lack of a parametric information needed in the case of AHT. A detailed description of the methods proposed and experimental results that demonstrate their success follow.

2. FACE DETECTION

As described in [1], skin-like regions in an image can be discriminated best by representing the image in the HSV color space and by choosing proper thresholds for the values of the hue and saturation parameters. Other similar color spaces that reside to the way the human vision interprets color information, like HSI or HLS, could also be used. Considering the oval-like shape of a face in general, it is convenient to search for the connected components of the skin-like regions using a region growing algorithm and to fit an ellipse to every connected component of nearly elliptical shape. This is a basic preprocessing step, because it restricts the search area for the algorithms that detect the several facial features, afterwards.

2.1. Parameteric feature extraction using AHT

In order to select an efficient set of distances, a parameterized representation of some facial features appears as a good approach towards the solution of the problem under study. This assumption holds for certain features that can be described in a realistic way by means of a geometric curve.

More specifically, the cheeks can be approximated by nearly vertical lines (with $\pm 20^0$ tolerance, with sign depending on which cheek we are referring to), and the chin by an upward parabola. A straightforward method to detect geometric curves in an image is the Hough transform or one of its variations [7].

The use of an Adaptive Hough Transform is proposed in [8]. It is a variant of the Hough transform proposed by Illingworth and Kittler [9], which utilizes a small fixed size accumulator and follows a *"coarse to fine"* accumulation until a desired precision for the values of the parameters is reached (or, equivalently, a desired resolution is achieved). This variation requires a reduced storage space compared to the standard transform due to the small accumulator size. Accordingly it also provides a better performance than the standard transform due to the reduced number of computations. The search for cheeks is based on the fact that a straight line is represented by an equation of the form:

$$y = mx + c \tag{1}$$

where m is the slope and c is the intercept of the line.

The accumulator array is thus two-dimensional, with each cell corresponding to a certain (m, c) pair of discrete values for the slope and intercept of the line. The inherent problem in equation (1) is that it cannot describe vertical lines, because $m \rightarrow \infty$. This can be solved by swapping the roles of x and y axes. A small accumulator size of 9×9 is used. A line segment is assumed to be interesting, if its length is at least 10 pixels in the edge image obtained by applying a Sobel edge operator on the original image [10]. The cheek corresponds to the dominant vertical line in the relevant subimage. Accordingly, the algorithm should search only for one line in this subimage. The main idea of the method is to "focus" on the parameter space on areas of large accumulation by performing iterations of the basic Hough algorithm and redefining the limits of the parameter ranges accordingly. This means that the search area in iteration i is

$$S^{(i)} = L_1^{(i)} \times L_2^{(i)} \times \ldots \times L_n^{(i)} \tag{2}$$

where $L_j^{(i)}$ is an interval (U_j, V_j) of possible values of the jth parameter a_j.

In our case, the relevant subimage is defined as a certain region of the largest best-fit ellipse found in the image. This ellipse is assumed to contain the main face region. The success of a search inside the ellipse is ensured by the fact that the ellipse possesses the same vertical symmetry as the face region that it contains. Let a and b be the magnitudes of the major and minor semiaxes and (x_0, y_0) be the center of the best-fit ellipse. Then a proper relevant subimage that gives satisfactory results is defined as a portion of the lower half of the ellipse. We choose

$$x_{min} = x_0 + 0.16 \cdot a \tag{3}$$

$$x_{max} = x_0 + 0.5 \cdot a. \tag{4}$$

If we search for a left cheek:

$$y_{min} = y_0 - 0.86 \cdot b \tag{5}$$

$$y_{max} = y_0 - 0.55 \cdot b. \tag{6}$$

If we search for a right cheek:

$$y_{min} = y_0 + 0.55 \cdot b \tag{7}$$

$$y_{max} = y_0 + 0.86 \cdot b. \tag{8}$$

The selection of this area of the ellipse ensures that the corresponding edge subimage contains part of the cheek and does not contain other significant features, like ears. Accordingly, the Hough process does not end up by selecting an undesirable result.

In the case of not nearly frontal face images, the above mentioned symmetry does not hold and correcting moves of the relevant subimage along the horizontal direction can be useful in order to compensate for rotations along the x-axis (see also [11]).

The chin can be described by the equation of an upward parabola:

$$x - x_v = g(y - y_v)^2 \tag{9}$$

where (x_v, y_v) is the vertex of the parabola and g is the parameter which controls how fast the parabola opens outwards. In this case, the accumulator array is three-dimensional. The size chosen in our implementation is $6 \times 6 \times 6$. A minimum segment length of at least 10 pixels is chosen in order to avoid an erroneous selection of a curve that contains isolated points of the edge image. The relevant subimage used for searching is defined as a certain region of the lower half of the best-fit ellipse as well. By exploiting the inherent symmetry, we prevent Hough transform to extract some other feature resembling to a parabola, like the lower edge of the lips.

The limits of the search area for the chin extraction are as follows:

$$x_{min} = x_0 + 0.62 \cdot a \tag{10}$$

$$x_{max} = x_0 + 0.85 \cdot a \tag{11}$$

$$y_{min} = y_0 - 0.53 \cdot b \tag{12}$$

$$y_{max} = y_0 + 0.53 \cdot b. \tag{13}$$

A compensation for rotations along the y-axis is provided by performing correcting moves along the vertical (major) axis of the ellipse.

2.2. Keypoint extraction using template matching

Unlike cheeks and chin, eyebrows do not possess a simple description by means of a geometric curve, because of their variations between different people. One aspect concerning eyebrows that is irrelevant to the specific shape of them is the upper point of the eyebrow. In fact, this implies that we have to detect the upper edge of the arch of the eyebrow. This edge constitutes a transition from a completely smooth area (the forehead) to an elongated approximately horizontal edge, as one is scanning the ellipse along the vertical axis. A template matching technique is more appropriate in this case. A prototype block has to be defined explicitly, simulating the form of the upper edge of an eyebrow, because no a priori information exists concerning the approximate position of the correct block (e.g. from a previous frame). Let P be the predefined block (acting as a mask) and B_i be the real block of the image under consideration, where $i \in S$ with S being the set of all blocks centered on the search area. The block that should be chosen as optimal is a block B_j of maximal correlation to the prototype block.

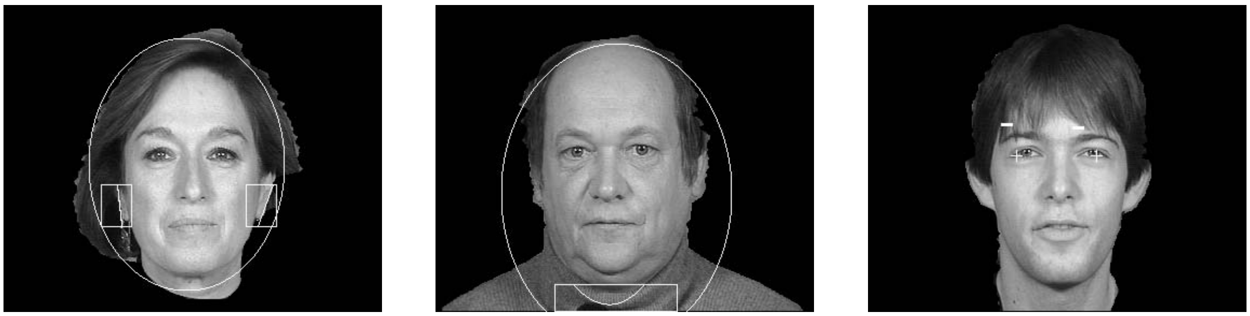
The size of the block to be searched for should be normalized according to the scaling factor of the face in the image, that is, how close the person was to the camera at the moment the shot was taken. A straightforward way to normalize the search process is to define the dimensions of the block based on the distance between the eyes. More specifically, a realistic consideration is to adapt a block of height equal to 0.125 of the eyes distance and of width equal to 0.5 of the eyes distance. The optimal block is expected to contain the upper edge of the eyebrow and can be used to provide a keypoint when defining the discriminating set of euclidean distances.

2.3. Snake deformation for face contour definition

In many cases, a problem arises when there is no sufficient edge information in the edge image so that the AHT algorithm can find the straight lines representing the cheeks or the parabola representing the chin. This is usually due to the poor lighting conditions during the time the shot is taken. A helpful technique to overcome

(a) (b) (c)

Figure 1. Examples of good results.



(a) (b) (c)

Figure 2. Examples of erroneous results.

this problem is to consider the best-fit ellipse surrounding the face as an active contour (snake) which can be deformed so its energy is minimized or, equivalently, its optimal position is determined. In general, the energy of a snake with snake elements (snaxels) $v_i = (x_i, y_i)$, $i = 0,..,M-1$, can be given by the equation:

$$E_{snake} = \sum_{i=0}^{M-1} E_{int}(v_i) + E_{ext}(v_i) \tag{14}$$

where E_{int} and E_{ext} denotes the interior and exterior energy terms of snaxels v_i.

The internal energy represents the forces acting on the contour and constraining it to be smooth. The first and second derivative of the function representing the contour are considered to comprise the internal energy term. Thus the equation for the internal energy can be written as [12]:

$$E_{int}(v_i) = w_1 \cdot \left| \frac{dv_i}{ds} \right|^2 + w_2 \cdot \left| \frac{d^2 v_i}{ds^2} \right|^2 \tag{15}$$

where s denotes the arc length along the ellipse. The first order term ensures the string-like behaviour of the snake, and the second order term ensures the rod-like behaviour. The extent to which a snake is allowed to be stretched or bent is controled by the weighting functions w_1 and w_2. Though these factors can be of fixed value, saving computation time, a better choice is to adapt them to the local characteristics of the snake elements, that is to compute them independently for each snaxel, based on the distance and curvature between adjacent snaxels.

The exterior energy refers to the forces that pull the snake to features on the image that are considered in each case to be significant. If, in our case, we consider the elliptical component to be the area which should attract the contour, then the external energy of the snake can be defined as:

$$E_{ext}(v_i) = -\sum_{(x,y)\in N_{int}(v_i)} 1 - c(x,y) + \sum_{(x,y)\in N_{ext}(v_i)} c(x,y) \tag{16}$$

where $c(x,y)$ stands for the indicator function of the pixels belonging to the skin-like region, and N_{int} and N_{ext} are the interior and exterior neighbourhoods of a snake element. Equation (16) implies that the parts of the region which are outside the contour should push the snake elements outwards, whereas the parts of the interior of the contour which do not belong to the face region should pull the snaxels inwards. Other significant features that could be considered contributing to the external forces acting on the contour, are edges, lines and terminations. A greedy algorithm can be used for minimizing the energy of the snake. This is based on the fact that the snake energy is minimal when the energy of each element is minimal (see [12]). Such an assumption might lead to unstable results, because of the localized action of the technique. This means that the contour might stuck at local minima. An obvious solution to this problem is to consider an area around the snaxel that is wide enough to contribute effectively to the power of the external forces. Other variations of a dynamic solution to the minimization of an active contour energy can be found in [13], [14] and [12].

3. EXPERIMENTAL RESULTS

The methods proposed above where tested on a set of 37 images of the M2VTS project multimodal face database. The results were rather satisfactory, considering the fact that the face images have not beeb really frontal in some cases.

Problems have been encountered in cheek-extraction when the false symmetry of the ellipse leads to bad definition of the relevant subimage, and thus to an erroneous extraction of some other feature considered as predominant. An example where ears are extracted instead of cheeks is shown in Figure 2(a), whereas Figure 1(a) shows a good result for both cheeks and chin, as well as the relevant subimages and the best-fit ellipse.

Similar problems cause the AHT to fail in the case of chin detection. The inability of the edge operator to detect weak edges caused by bad luminance, is more obvious here. An example of extraction of some other feature instead of the chin is given in Figure 2(b).

The correct extraction of the eyebrows is heavily dependent on the position of the eyes, as one would expect. If at least one eye is detected at a wrong position, and especially when the distance between the eyes is bigger than the actual one, the size of the block to be matched is incorrect. This leads to the false extraction of another feature. Even when eyes are correctly extracted, a problem may arise if hair covers the forehead and the eyebrows, as is shown in Figure 2(c). A correct result is shown in Figure 1(b).

Table 1 shows the results of extraction of cheeks, chin and eyebrows for the set of images mentioned above:

features	detected [%]	correctly [%]	falsely [%]
cheeks	86	81	19
chin	84	71	29
eyebrows	84	66	34

Table 1. Detection rates for facial feature extraction

The rate of correct detection of eyebrows in the table is rather low because, as stated above, the correlation between eyes and eyebrows is very high, so false results for the eyes also lead to false results for the eyebrows. If we consider the set of images where a correct detection of eyes has been performed, the rate of correct eyebrows detection increases to about 95%.

In cases of images that suffer from a lack of edge content, the ellipse deformation algorithm succeeds in providing additional information for the later definition of feature distances. This is illustrated in Figure 1(c). It fails only in extreme cases, when the main connected component of the image contains features that deform the real face region dramatically (e.g. when hair surrounding the oval face region are included in the connected component).

4. CONCLUSIONS

A combination of methods based on adaptive Hough transform, block matching and active contour deformation variations are proposed in order to provide sufficient feature information to be used in the definition of a set of geometric distances for face recognition purposes. Examples denoting the success of the different methods, as well as their failure in certain cases, are provided together with a justification of the results obtained.

ACKNOWLEDGEMENTS

This work has been carried out within the framework of the European ACTS-M2VTS project.

REFERENCES

[1] K. Sobottka and I. Pitas, "Face localization and facial feature extraction based on shape and colour information", in *Proc. of the IEEE Int. Conf. on Image Processing (ICIP-96)*, pp. III-483–III-486, Lausanne, Switzerland, September 1996.

[2] G. Yang and T.S. Huang, "Human face detection in a complex background", *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.

[3] C. Kotropoulos and I. Pitas, "Rule-based face detection in frontal views", in *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASP 97)*, Munich, Germany, April 21-24, 1997, to appear.

[4] R. Chellapa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

[5] R. Brunelli and T. Poggio, "Face recognition: Features versus Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, October 1993.

[6] K. Sobottka and I. Pitas, "A fully automatic approach to facial feature detection and tracking", in *Lecture Notes on Computer Science: First International Conference on Audio- and Video-based Biometric Authentication (AVBPA 97)*, Crans-Montana, Switzerland, March 12-14, 1997, to appear.

[7] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. I. Reading, Massachussets, Addison-Wesley, 1992.

[8] X. Li and N. Roeder, "Face contour extraction from front-view images", *Pattern Recognition*, vol. 28, no. 8, pp. 1167-1179, 1995.

[9] J. Illingworth and J. Kittler, "The Adaptive Hough Transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 690–698, September 1987.

[10] I. Pitas, *Digital Image Processing Algorithms*, Prentice Hall, UK, 1993.

[11] S. Fischer, B. Duc and J. Bigün, "Shape normalization for face recognition", in *Lecture Notes on Computer Science: First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 97)*, Crans-Montana, Switzerland, March 12-14, 1997, to appear.

[12] D. J. Williams and M. Shah, "A Fast Algorithm for Active Contours and Curvature Estimation", *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 55, no. 1, pp. 14–26, January 1992.

[13] K. M. Lam and H. Yan, "An Improved Method for Locating and Extracting the Eye in Human Face Images", in *Proc. of Int. Conf. on Pattern Recognition (ICPR '96)*, pp. 411–415, Vienna, Austria, August 1996.

[14] S. R. Gunn and M. S. Nixon, "Snake Head Boundary Extraction Using Global and Local Energy Minimisation", in *Proc. of Int. Conf. on Pattern Recognition (ICPR '96)*, pp. 581–585, Vienna, Austria, August 1996.