AN AUTOMATIC ALGORITHM FOR FACIAL FEATURE EXTRACTION IN VIDEO APPLICATIONS

Y. Sheng, A. H. Sadka, A. M. Kondoz

Centre for Communication Systems Research
University of Surrey, Guildford, GU2 7XH, U.K.
Email: Y.Sheng@surrey.ac.uk

ABSTRACT

An enormous market demand for spreading applications such as videoconferencing and videophony, has led to considerable investigation of facial feature extraction. Although an avalanche of algorithms have been proposed for facial feature extraction, due to some unpredictable and objectively existing factors like lighting condition, partial occlusion, as well as the variability of face expression, automatic facial feature extraction still remains a challenging topic, seeking more robust solutions. Moreover, automation has a strict demand in performance consistency of each part of a system, so as to be capable of smoothly and robustly coping with various situations. In other words, automation highly needs operators to supervise the system as less as they can, that is, to avoid the use of the predefined threshold values and training data. This paper addresses an automatic facial feature extraction algorithm, by which facial features such as eyes, eyebrows, mouth and chin can be accurately and rapidly extracted.

1. INTRODUCTION

Strictly speaking, automatic facial feature extraction is comprised of two phases. A likely human face is first detected from the unknown background in a head-and-shoulder video frame, and then facial features such as the eyes, eyebrows, mouth and chin are in turn extracted on the detected face region. Detail of the existing methods in face detection is findable in a recent survey [1]. For the proposed algorithm in this paper, the face detection method addressed in [2] is employed.

As for facial feature extraction, various approaches aimed at different sets of facial features have been proposed. The simplest means to find facial features is based on the fact that brightness of the eyes and mouth is lower than the rest of the face. In [3], the topographic grey level relief in the horizontal and vertical directions is analysed, respectively, while in [4] the irises are located through a region-growing search. However, these methods are only capable of finding the approximate position of facial features. Active Contour Model, i.e. snake, is an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges [5]. The snake is renowned for its flexibility of fitting object boundary, but it has a disadvantage of easily converging into the local minimum. Deformable templates, the most widely used method for boundary location of facial features applied to the eyes, mouth, chin contour and even nose, have been suggested for decades [6]. In this method, *a priori* elastic parabolic curve models are set up for the eyes, mouth and so on. These facial features are defined in terms of edges, peaks and valleys corresponding to different energy functions. The best-fit elastic model is found by minimizing an energy function of the edges, peaks and valleys. One shortcoming is that the deformable template must be initialised in the proximity of the facial features [1].

Taking the above points into account, based on face detection results from [2], a novel, automatic and robust facial feature extraction algorithm is addressed in this paper. There are few assumptions made before we go ahead. Firstly, the processed images or video frames must be the type of the head-and-shoulders that normally comes in real-time videophony and videoconferencing applications. Secondly, no occlusion appears on the face, i.e. no facial feature covered. The rest of the paper is structured as follows: the proposed facial feature extraction algorithm is described in Section 2; Section 3 formulises some relevant knowledge; Section 4 demonstrates the feasibility of the proposed algorithm through some experimental results. Finally, conclusion is drawn in section 5.

2. FACE FEATURE EXTRACTION

Facial Features connotate some evident characteristics of human face, such as eye corners, mouth boundary etc. In our algorithm, not only feature points, but also those feature boundaries are taken into account. In this section, the proposed methods of extracting mouth, eyes, eyebrows and chin are in turn stated as follows.

2.1. Mouth Contour Extraction

Using deformable templates to extract mouth contour is increasingly getting popular because of the continuity and flexibility of the deformable templates. The deformable templates can automatically adhere to feature boundaries by adjusting their size, orientation in terms of *a priori* elastic parabolic curve models. However, there are still some bottlenecks, seeking more robust solution. The current hotspot issues in using deformable templates for mouth contour extraction can be focused into the following categories:

- Template position initialisation. Before any deformation, the mouth template must be initialised in the vicinity of the mouth. This involves automatic detection of the mouth.
- Criterion of mouth being open or closed. This is to decide which template is to be used. However, uncertainty of contents inside the mouth, such as presence of teeth, raises the difficulty of discrimination.
- Selection of strength constraints. It seems that different researchers prefer different strength constraints.
- How to deform the template. After mouth template positioning, the problem of how to deform the template seeking the best-fit shape emerges.
- Accuracy of template. Sometimes mouth contours are not smooth parabolas, leading to mismatch.

Considering the points above, a mouth contour extraction algorithm with a staged deformable template is proposed. To position the template, the mouth corners are first extracted by using the method addressed in [7]. The search area can be narrowed into the lower half of the face. Chrominance component Cr is adopted to segment mouth because of its sensitivity to red colour. For the sake of algorithm automation, a dynamic threshold value can be evaluated by an equation, $T = C_{max}(i, j) \times e$, where C_{max} symbolises a local chrominance peak of the search area, e is an elastic coefficient, empirically set to 0.934.

The proposed staged deformable template is adapted onto the mouth, following a criterion of mouth being open or not, first introduced in [7]. As the name of the proposed method implies, middle points of the inner lip contour, the inner lip contour and the outer lip contour are sequentially located. [7] also proposed an algorithm of extracting the mouth corners and the middle points of the inner lip contour. Based on these four points, two parabolic templates are placed on the inner upper and lower lip contours. Positioning of two parabolic curves fixes the inner part of the entire mouth template so that only outer part of the mouth contour is required to be located. The edge and region strength are taken into account in locating mouth contour. To reduce the impact of the teeth, the outer lip contour template is applied to the Cr image. Compared with other methods, an advantage of the proposed algorithm is gradual extraction of the mouth features. There are only two strength functions considered as deforming the templates. It considerably reduces the computational complexity, and improves the degree of accuracy.

2.2. Boundary Location of Eyes and Irises

Since the deformable templates are some strength-dependent curves, they have a high demand in pattern articulation. Therefore, the deformable templates are not robustly applicable to those images with very low resolution, e.g. the images in QCIF format. Considering the above reason, we only employ the concept of the parabola rather than the deformable templates. The method of eye boundary extraction is described as follows.

The search area can be first narrowed into the upper half of the face. From observation, the texture around and inside the eyes is complicated. This makes the eyes differ from the peripheral areas on the face. With this revealing idea in mind, an edge-oriented method is adopted, followed by morphological processing. [7] gives a detailed description of this method, by which two corners as well as the middle points of the upper and lower eyelids of each eye can be robustly extracted.

Regardless of the case that eyes are closed, the edge of the iris can be viewed as a circle embedded within the eye. Therefore, the method of extracting irises is based on the use of Hough transform with edge detection.

2.3. Eyebrow Extraction

The search area of the eyebrows can be narrowed by the fact that the eyebrows are permanently over their corresponding eyes. In combination with low intensity of the eyebrows and a system demand of automation, an automatic eyebrow segmentation approach is summarised below. Let $I_{i,j}$ denote the intensity of a certain pixel within the narrowed search area, where the subscripts i and j indicate the vertical and horizontal coordinates of the image plane, respectively. The downward intensity gradient at $I_{i,j}$ can be formulised as $g_{i,j} = I_{i,j} - I_{i,,j+1}$. Then we seek the maximum gradient value G_j on the current column because the place where the maximum gradient arises must be on the border of the eyebrows. Once the boundary position of the current column is discovered, we can determine the threshold value of the column by averaging the intensities of the two adjacent pixels. To normalise the threshold, the mean of the determined thresholds on all the columns of the eyebrows is finally calculated. The determined threshold value is used to segment the eyebrows.

2.4. Chin Extraction

The chin contour can be viewed as a parabola with two end points sited on the intersections of, the edge of the detected face and the line passing through two mouth corners. So a parabolic template is used to locate the most likely chin contour, based on the edge image.

3. FORMULATION OF ALGORITHM

3.1. Coordinate Transformation Modeling

In *Fig 1*, the axis J and I define the coordinate of an image plane. The coordinate system YOX is an arbitrary 2D subspace in the image plane, where $O(i, j)$ indicates the origin. Rotate YOX clockwise until the axis Y is in parallel with the image axis I, forming a new coordinate plane, yOx. The clockwise rotation angle, Xox, is symbolised by θ. Assume there is an arbitrary point $P(X, Y)$ in the plane YOX. The coordinate transformation of P into the plane yOx can be written as:

$$x = X\cos\theta + Y\sin\theta \qquad y = Y\cos\theta - X\sin\theta \qquad (1)$$

where (x, y) is the coordinate of P in the plane yOx. Meanwhile, the coordinate transformation functions between the image plane and the plane YOX can be summarised by:

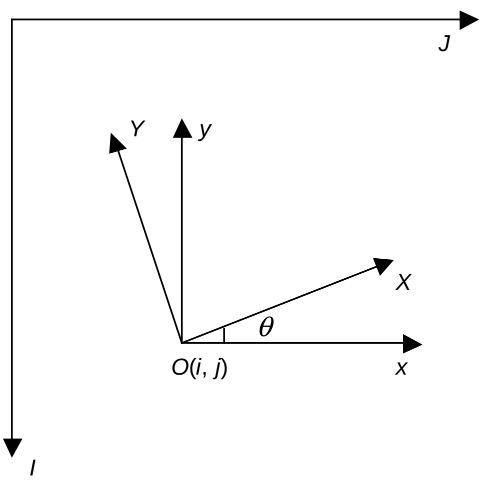

Fig 1 Coordinate transformation modeling

$$I = i - y \qquad J = j + x \qquad (2)$$

Thus, substitute equation (1) into (2), giving an expression of the point $P(X, Y)$ in the image plane:

$$I = i - Y\cos\theta + X\sin\theta \qquad J = j + X\cos\theta + Y\sin\theta \qquad (3)$$

3.2. Deformable Templates

Regardless of rotation, the equation of a parabola is given by:

$$Y = H \times \left(1 - \left(\frac{X}{L}\right)^2\right) \qquad (4)$$

As shown in *Fig 2*, H and $2L$ symbolise the high and width of the parabola, respectively. Combining equation (4) with (3) yields:

$$I = i - H \times \left(1 - \left(\frac{X}{L}\right)^2\right) \times \cos\theta + X\sin\theta$$

$$J = j + X\cos\theta + H \times \left(1 - \left(\frac{X}{L}\right)^2\right) \times \sin\theta \qquad (5)$$

Equation (5) stands for the expression of a parabolic template in the image place.

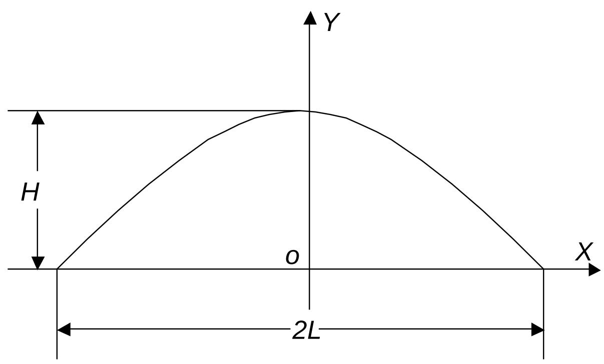

Fig 2 Parabola template

The edge strength E_e of the deformable templates is given by:

$$E_e = -\frac{1}{2L} \int_{2L} e(x, y)ds \qquad (6)$$

where $2L$ is the length of arc (see *Fig 2*); $e(x, y)$ denotes the edge intensity; s indicates the trajectory of facial feature contours. Calculation of the region strength is based on Cr defined as:

$$E_r = -\frac{1}{A} \int\int_A C(x, y)dr \qquad (7)$$

where A stands for the area of the regions, $C(x, y)$ indicates chrominance intensity; r represents the scanned area. In our proposed algorithm, the mouth contour extraction involves both the edge and region strengths, while the chin contour extraction only employs the edge strength. By minimizing these strength constraints above, the deformable templates can be automatically adhered onto the feature boundaries.

4. EXPERIMENTAL RESULTS

Various video sequences have been tested using the proposed facial feature extraction algorithm. Two of them are shown in *Fig 3*. These images involve the cases when the mouth is open and closed, the head is tilting, and the expressions of joy, surprise. As can be seen, four points, i.e. the corners and the middle upper and lower points on the eyebrows, contours of the eyes, irises, mouth lips, and chin are robustly extracted.

5. CONCLUSION

This paper presents an automatic facial feature extraction algorithm aimed at video applications. A staged deformable template for mouth feature extraction, as well as some improved methods to extract features of the eyes, irises, eyebrows and chin are proposed. The feasibility of the proposed facial feature extraction algorithm has been successfully demonstrated, showing that its robustness is highly improved, and that it can robustly track the facial features.

6. REFERENCES

[1] M.H. Yang, D.J. Kriegman and N. Ahuja, “Detecting faces in images: a survey”, IEEE Trans. on PAMI, Vol.24, No.1, pp.34-58, January 2002.

[2] Y. Sheng, A.H. Sadka, A. Kondoz, “Automatic face segmentation for 3-d model-based video coding”, IEE International Conference on Visual Information Engineering, UK, pp.274-277, July 2003.

[3] K.Sobottka and E.Pitas, “Face localization and feature extraction based on shape and color information”, Proceedings IEEE International Conference, Image Processing, pp.483-486, 1996.

[4] B. Cao, S. Shan, W. Gao, D.Zhao “Localizing the iris center by region growing search”, Int. Conf. on Multimedia and Expo, Vol.2, pp.129-132, 2002.

[5] M. Kass, A. Witkin, D. Terzopoulos, “Snake: active contour models”, Int. Journal of Computer Vision, pp.321-331, 1988.

[6] A.L. Yuille, P.W. Hallinan and D.S. Cohen, “Feature extraction from face using deformable templates”, International Journal of Computer Vision, Vol.8, No.2, pp.99-111, 1992.

[7] Y. Sheng, A.H. Sadka, A. Kondoz, “Fast and automatic facial feature extraction for 3-d model-based video coding”, proceedings of 4th WIAMIS, Queen Mary, Universityof London, pp. 387-390, April 2003.

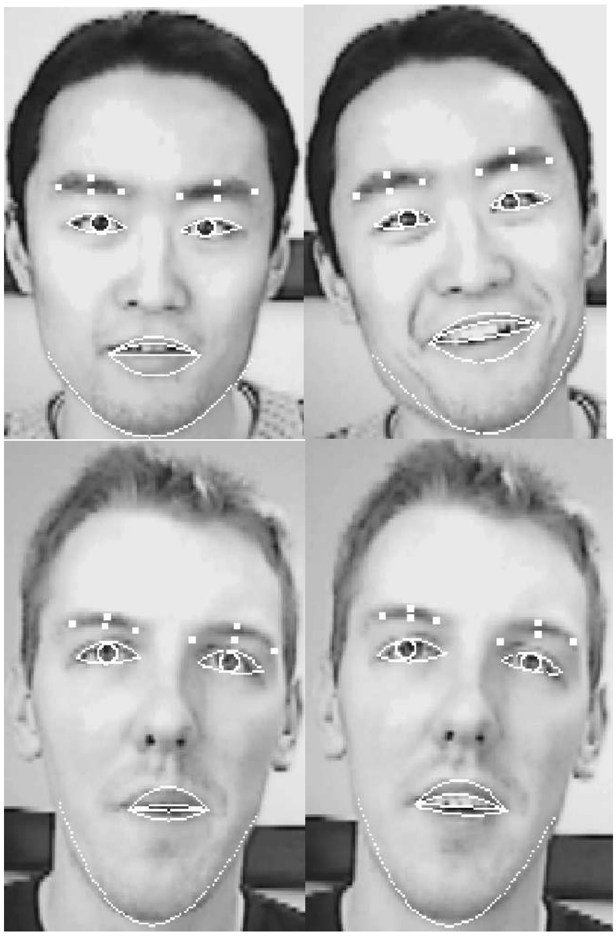

Fig 3 Extracted facial features