NOSE SHAPE ESTIMATION AND TRACKING FOR MODEL-BASED CODING

Lijun Yin and Anup Basu

Department of Computing Science, University of Alberta, Edmonton, Alberta, T6G 2H1, Canada.

ABSTRACT

Feature extraction on the face plays an important role in applications of model based coding and human face recognition. Traditionally, eyes and mouth are considered to be the most significant features contributing to different facial expressions. However, detecting and tracking the nose shape is non-trivial, and plays an equally important role as eyes and mouth for model based coding, especially for analysis and synthesis of *realistic* facial expressions. In this paper, a feature detection method on the facial organ areas is presented. Individual templates are designed for the nostril and nose-side. First, the feature regions are limited to certain areas by using two-stage region growing methods. Second, the pre-defined templates are applied to extract the shape of the nostril and nose-side. Finally, the extracted feature shapes are exploited to guide a facial model to complete an accurate adaptation. The advantage of the proposed scheme is demonstrated by experiments on real video sequences for low bit rate video coding.

1. INTRODUCTION

Facial feature extraction is critical for facial expression analysis and synthesis, face recognition, and model adaptation for model based coding. Most researchers have been focusing on the extraction of the eye and mouth shapes in the past decades [1, 6, 7, 3, 4]; few have paid attention to nose shape, which is an important part of facial expression recognition and generation. One of the important application of extracting and tracking nose shape is to generate a realistic facial expression in the scenario of SNHC (Synthetic-Natural Hybrid Coding) formed within MPEG4 [2]. The important nose features lie in the shapes of nostril and nose side. The accurate detection of nose shape can be important for facial expression synthesis, especially when a person is smiling or laughing. Subtle change of nose shape results in a different expression, even a different emotion, of a person. In this paper, we propose an efficient method to detect and track the nostril shape and nose-side shape based on individual deformable templates. First, the silhouette of the moving head is extracted. In the extracted head area, facial feature regions can be localized by using a two-stage region growing method. Then, the feature shape of a nose can be extracted in the limited feature regions using pre-defined deformable templates. In Section 2, an algorithm for localizing facial feature regions is presented. Section 3 and 4 describe how to extract the nose shape, and how to adapt the model on to a face. Experimental results are presented in Section 5. Finally, concluding remarks are given in Section 6.

This work is supported in part by the Canadian Natural Sciences and Engineering Research Council. Email: {lijun, anup}@cs.ualberta.ca.

2. FEATURE REGIONS LOCATION

In order to track the head motion, an active tracking algorithm for tracking the head silhouette is developed in [5]. The head silhouette greatly reduces the feature search area. Since facial skin other than facial features and hair exhibits a similar property in color and luminance, region growing is suitable for extracting the connected skin area. In order to make the facial feature detection less sensitive to noise, a two-stage region growing algorithm is developed – global region growing and local region growing. In global region growing, a large growing threshold is selected in order to explore more skin area. Only the most distinct feature locations, such as the darkest parts of the eyes, the nostrils and in-between the lips, remain, while their size information is not kept. The second stage is local region growing, where the region information is estimated in the individual feature areas. The initial seed pixel of the skin is selected in the local area, and a smaller growing threshold can be selected so that a small skin area surrounding the facial organ is detected, and produces as much feature information as possible.

2.1. Feature center location by global region growing

The facial features are searched within the obtained head silhouette area. In general lighting conditions, the hair and the face exhibit distinct color attributes. Also, the facial features of the iris, eyebrow, in-between lips, and nostril have a distinguishable darkness compared to smooth facial skin, which usually exhibits uniformity in color. Hence, the feature regions can be extracted by segmenting the image using region growing based on the color components, *i.e.*, luminance (*Y*) and chrominance (*U*,*V*). The image is partitioned based on the checker-board distance in region growing, the distance D between seed pixel (y_s, u_s, v_s) and the growing pixel (y_i, u_i, v_i) is defined in Formula 1. Larger variations are allowed in the luminance component of a region than in the chrominance component, in order to ignore small changes in luminance caused by shading. Therefore weights (1/4,1,1) are used for YUV. Regions smaller than a certain number of pixels are not taken into consideration, and are removed.

$$D = max(\frac{1}{4}|y_i - y_s|, |u_i - u_s|, |v_i - v_s|) \qquad (1)$$

The region growing is performed by selecting the central point of the head area as a seed pixel. The size of grown region must be examined to ensure that it is a reasonable face region; if it exceeds a predefined range, a new seed is selected from the neighboring pixels, or the threshold is adjusted, to generate a new region. This process is iterated until a skin region is found. After the skin region growing, a number of regions (blobs) are obtained. In order to extract the facial organ blobs (*e.g.*, eyes, mouth, nose, etc.), the top blobs and the bottom blobs (*e.g.*, hair, collar, cloth, etc.) are removed. The feature blobs are then processed for further classification.

Classification of the regions of eyes, nose and mouth

After obtaining the set of feature blobs, the next task is to classify them into four groups, *i.e.*, left eye blobs, right eye blobs, nose blobs and mouth blobs. These four classes are distinguishable by the vertical distance and horizontal distance of the blobs. The k-means classifier is an ideal tool to classify them. Figure 1 illustrates feature blobs classification and feature center estimation. To explain the clustering in Figure 1, several terms are described as follows: (i) **Central Point** of a blob is the pixel with an average coordinate of the blob. (ii) **Blob Distance** is the distance between the central points of two blobs. *Vertical Distance* is the distance of two blobs along the *y*-axis. *Horizontal Distance* is the distance of two blobs along the *x*-axis. (iii) **Vertical clustering** refers to the class classification using vertical distance. **Horizontal clustering** refers to the class classification using horizontal distance. After obtaining the four feature classes, the *class center* for each feature class can be obtained by taking the average of all blob centers of that class. Figure 2 (col 1) shows some examples of the feature estimation.

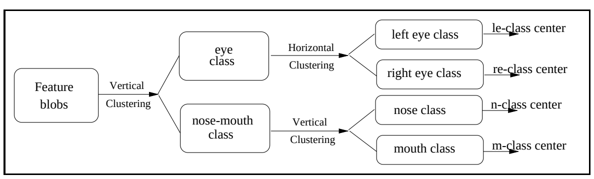

Figure 1: Feature region classification and center estimation

2.2. Location of feature regions by local region growing

The local area of each feature (organ) is determined by the estimated location of the feature center and the predefined surrounding size. To grow a skin region in the local area, a seed pixel is selected from the periphery skin area of the feature organ. A smaller growing threshold D_{T_S} is predefined. The feature blobs with large area are created, based only on the luminance of the image rather than the three color components. The feature region is a complementary set of the grown region within a surrounding window. Finally the region size of each organ (called the *organ window*) can be derived from the extracted feature blobs which are represented by a binary image. The feature area is represented by a high value while the background area is represented by value zero, as shown in Figure 2 (col 2). The size of the detected feature organs is determined by the pixels of the left-most, right-most, top and bottom of the feature region. By using local region growing, feature areas with clear boundaries can be obtained, as shown in Figure 2 (col 3).

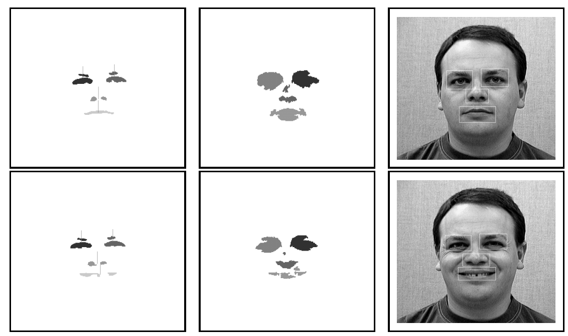

Figure 2: Left: Classified features after global growing (the bottom terminal of each line is the center of each feature); Middle: Extracted feature blobs after local growing; Right: The restricted window area.

3. SHAPE ESTIMATION

The above work shows how facial features are restricted in the corresponding windows. In this section we address the issue of extraction of the real shapes for the model adaptation. The shape of facial features mainly refers to the contour of eye, eyebrow, mouth, nose (nostril and nose side), and chin. A variety of approaches have been proposed for the shape detection of facial features, including deformable template matching [7], Hough transforms, and color image processing [4]. The deformable template method requires an accurate initial localization, it also inclines to shrink. Many energy terms make the computation more complex. We overcome some of these difficulties by improving the initial localization process by making full use of the color information. Eye detection uses Hough transform to search the iris position and size in order to determine the initial location of the eye template. The work on eye detection using saturation information has been developed in our previous work [6]. Here we focus on the description of nose feature detection.

3.1. Detection and tracking of nose shape

The important nose features lie in the shapes of nostril and nose side. The accurate shape of the nostril and the nose side can be important for composing a facial expression, especially when a person is smiling or laughing, the nostril shape and the nose side are obviously changed.

Nostril estimation

The nostril has a distinctive darkness from the facial skin. As mentioned in the previous section, color-based region growing can roughly detect the position and the approximate shape of the nostrils. In order to detect the nostril shape correctly, a geometric template is further applied on the nostril region, which is a twisted pair curve with a leaf-like shape, as shown in Figure 3. The nostril template is defined as a part of a twisted pair curve, which is represented in the polar coordinate system as:

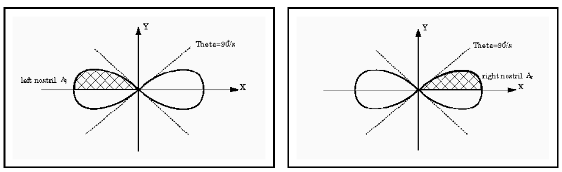

Figure 3: Template of nostrils (Left and Right)

$$\rho^2 = a^2 cos(s\theta) \tag{2}$$

The nostril on the right side and the left side of a nose can be represented by the up-right curve and the up-left curve, respectively. Parameter a is the width of a nostril. Parameter s controls the shape of the curve, which can take a real value in the range [1,10]. The smaller the s value, the thicker the leaf-shape appears. Besides the parameters a and s, the orientation of the x axis and the position of the origin o can also be adjustable. The relationship between (x, y) in Cartesian coordinate system and (ρ, θ) in polar coordinate system is defined as:

$$x = \rho cos(\theta), y = \rho sin(\theta) \tag{3}$$

From the detection of the nose regions in the previous section using color-based region growing, the initial width of the nostril can be determined by calculating the distance between the left-most pixel and the right-most pixel in the nostril region. The initial orientation of the nostril can be determined by the eigenvector with the largest eigenvalue of the nostril pixel region. The initial value of the shape control parameter s is set to 2. Note that the two nostrils are assumed to be symmetric with respect to the center line of the face; a nostril missed in the previous stage can be by flipping a detected one. After the initial parameters are estimated, the deformable template matching can be performed. The cost function for energy minimization is defined based on the assumption that: (1) the nostrils have distinctive darkness compared to the skin, and (2) the luminance gradient has a high value at the border of a nostril. The cost function for the right nostril is defined as:

$$E^{nl} = k_1 E^{nl}_{lumin} + k_2 E^{nl}_{grad} \tag{4}$$

where

$$E^{nl}_{lumin} = \frac{1}{|A_r|} \int\limits_{A_r} \Phi_{lumin}(\vec{x}) dA \tag{5}$$

$$E^{nl}_{grad} = -\frac{1}{|B_r|} \int\limits_{B_r} \Phi_{grad}(\vec{x}) ds \tag{6}$$

The above equations estimate the brightness of the nostrils, and the luminance gradient, respectively. A_r is the area inside the upper-half of the right leaf-shape curve. $\Phi_{lumin}(\vec{x})$ is the value of luminance component of the nostril region. $\Phi_{grad}(\vec{x})$ is the edge magnitude. The cost function for the left nostril has the same definition as the one for the right nostril, except the left leaf-shape curve is used instead. During the minimization every parameter of the deformable template (location, orientation, width, shape) can be changed.

Nose side estimation

The nose side has a shape like a vertical parabola. (Assuming that the nose orientation has been normalized into the straight direction which is parallel to the face center line.) A pair of templates of the nose side (both the right side and the left side) are defined in Figure 4,

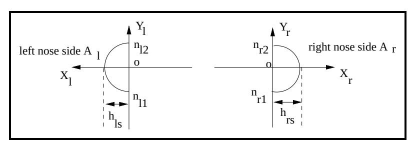

Figure 4: Template of nose side

$$x_l = (h_{ls}(1 - (\frac{y_l}{|n_{l2} - n_{l1}|})^2), x_r = (h_{rs}(1 - (\frac{y_r}{|n_{r2} - n_{r1}|})^2) \tag{7}$$

The initial corner of the nose side n_{s1} is fixed since the nostril has been estimated. The parameter n_{s2} and the width h_{ns} need to be estimated by minimizing the energy, as defined below. The cost function for the right side of a nose is:

$$E^{ns} = k_1 E^{ns}_{lumin} + k_2 E^{ns}_{grad} \tag{8}$$

where

$$E^{ns}_{lumin} = -\frac{1}{|A_r|} \int\limits_{A_r} \Phi_{lumin}(\vec{x}) dA \tag{9}$$

$$E^{nl}_{grad} = -\frac{1}{|B_r|} \int\limits_{B_r} \Phi_{grad}(\vec{x}) ds \tag{10}$$

A_r is the area inside the parabola, which has bright luminance value $\Phi_{lumin}(\vec{x})$ in the fold on the side of the nose. $\Phi_{edge}(\vec{x})$ is the edge magnitude. The cost function for the left side of a nose has the same definition as the one for the right side. Figure 5 shows some sample frames with the detected nostrils and nose sides.

Figure 5: Deformable template matching on nose shape.

4. FACIAL MODEL TRACKING

After the facial feature regions and shapes on various organs are estimated, a 3D wireframe model can be matched onto the individual face to track the motion of facial expressions using energy minimization (see [6] for the detailed model adaptation algorithm). Fig. 6 shows some sample results of model tracking.

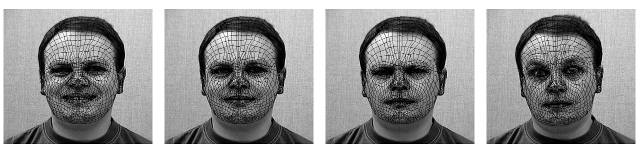

Figure 6: Model tracking ("Mario": frame 20, 32, 42, 62).

5. EXPERIMENTAL RESULTS

We use a camera mounted on an active platform (pan/tilt) to take an active video sequence, which shows a talking person with an unconstrained background. The camera rotation is less than $\pm 5°$. Note that our deformable template matching algorithm only uses several single steps to fit the template to the facial features, it avoids the computational complexity for a large amount of search of the parameters, and the template shrinking problem. Figure 7 shows some samples of feature shape detection. Combined with our previous work

Figure 7: Top: coarse regions by two-stage region growing; Bottom: detected facial features (*Guan*: frame 10, 24, 41)

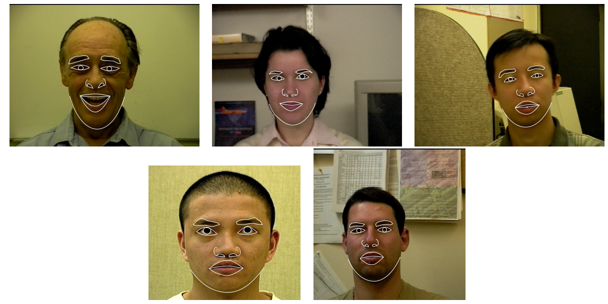

Figure 8: Detected features on video [top]:*Stan, Dana and Xun*; [bottom]: *Alau, Dima*

[6, 5], 5 other video sequences are tested as well. Notice that nostril and nose-side are among the features detected in Figure 8. The overall performance (measured over all tested frames) gives an indication of the capability of the system to detect most nose features correctly (only 18 frames out of 270 frames showed that the position error in nose area was beyond 3 pixels).

6. CONCLUSION

In this paper we proposed a nose shape detection method for low bit rate model based coding. This new method limits the search region, and uses the shape-adaptive model as a remedy for the shrinking effect of deformable templates. Initial results show that this approach is feasible in practical applications. In our future work, more robust feature detection algorithms (to deal with large rotation of a face) and real time issues will be investigated.

7. REFERENCES

[1] G. Chow and X. Li. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26(12):1739–1755, December 1993.

[2] H.H. Chen. *et al*. Special issue on synthetic/natural hybrid video coding. *IEEE Trans. CAS for Video Tech.*, 9(2), 1999.

[3] C. Lin and J. Wu. Automatic facial feature extraction by genetic algorithm. *IEEE Transactions on Image Processing*, 8(6):834–845, June 1999.

[4] H. Wu, T. Yokoyama, D. Pramadihanto, and M. Yachida. Face and facial feature extraction from color image. In *ICAFGR'96, Killington, Vermont, USA*, p345-350, Oct. 1996.

[5] L. Yin and A. Basu. Integrating active face tracking with model based coding. *Pattern Recog. Letters*, 20(6), 1999.

[6] L. Yin and A. Basu. Realistic animation using extended adaptive mesh for model based coding. In *EMMCVPR'99, Springer-Verlag Lecture Notes Series. York, UK.*, July 1999.

[7] A.L. Yuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, 1992.