Available online at www.sciencedirect.com
SCIENCE DIRECT®
Engineering Applications of Artificial Intelligence 19 (2006) 545–555
Engineering Applications of
ARTIFICIAL INTELLIGENCE
www.elsevier.com/locate/engappai

# Multi-view based face chin contour extraction

Xinliang Ge*, Jie Yang, Zhonglong Zheng, Feng Li

*Institute of image processing and pattern recognition, Shanghai JiaoTong University, Shanghai 200030, China*

Received 5 January 2005; received in revised form 7 August 2005; accepted 28 December 2005
Available online 20 March 2006

**Abstract**

Chin contour is an important facial feature to build a 3D morphable model, the core step of which is to establish feature points correspondence between each face in the training set and the reference face. In this paper, robust face detection is implemented firstly using probabilistic method. A probability of detection is obtained for each image of different position and at several scales and poses. Then, the chin contours are extracted accurately using the active shape model (ASM), which depends on the parameters obtained from the face detection. From frontal (0°) to profile (90°) faces that are equally divided into 10 parts, we train 10 flexible models. Then, different flexible models are used to extract the face chin contour according to the corresponding face pose. Experimental results show that the proposed approach can extract the chin contours of different people across different poses with good accuracy.


*Keywords:* Chin contour extraction; Pose estimation; Active shape model

## 1. Introduction

Reliable and concise face detection as well as features extraction plays an important role in many applications, such as human face recognition and very low bit-rate video coding, etc. In different applications, the operating conditions and the required performance characteristics of the face recognition system are often different, such as light conditions and face poses, which may exhibit greater changes than the differences between different people's images. The variations of illumination and pose are the main challenge for face recognition nowadays. The following steps are indispensable to resolve these problems. First, faces are located in the image using some face detection approaches such as SVM-based methods (Lie et al., 2001) and Adaboost based methods (Yongmin et al., 2000). Second, face recognition is performed based on the results of face detection. However the existing methods (Kumar et al., 2003; Perlibakas, 2004; Othman and Aboulnasm, 2003) are confronted with the austere challenges mentioned above, and the face recognition ratio is very low because of the changes of rotation in the depth direction. Therefore, new approaches and new ideas are proposed to deal with the problem of depth rotation. The three-dimension face recognition can deal with these difficulties well, although it is in the preliminary research stage in face recognition. 3D face recognition system (Koh et al., 1999) firstly builds a morphable model, and its core step is to establish feature points correspondence between each face in the training set and a reference face. Chin contour is an important part of facial features because it is the dividing line between the face and neck. Therefore it is absolutely necessary to extract chin contour in face recognition applications. However it is difficult to extract chin contour due to its weak contrast with neck. Generally speaking, different chin contours extraction algorithms are based on the deformable template (Yang et al., 2002; Yuille et al., 1992) and the statistical model (Feraud et al., 2001; Liu, 2004). The advantage of the former strategy is simple, some geometrical shape models represent eyes, mouth and chin contour, etc. However, invariable graphics cannot build the best model for all kinds of eye, mouth, etc., because they do not adapt to the changes of the illumination, facial pose, and facial expression. Therefore, the approach is difficult to extract the facial features robustly. The latter can overcome the disadvantages mentioned above. Active shape model (ASM), active

*Corresponding author. Tel.: +86 21 62933739; fax: +86 21 62932035.
*E-mail address:* xinliang@sjtu.edu.cn (X. Ge).

0952-1976/$ - see front matter 
doi:10.1016/j.engappai.2005.12.010

appearance model (AAM) and direct appearance model (DAM) are the typical statistical model approaches. The search performances of ASM have stronger superiority than the other two. The main drawback of ASM is its sensitivity to the flexible model initialization position. However, ASM can gain them on the basis of the face detection.

*Phase I-face detection*. This phase detects the regions of interest that potentially contain faces. A detection probability is associated closely with each pixel, furthermore for different scales and two different views, i.e. frontal and profile faces (Volker and Thomas, 2003). The parameters of face position, scale and pose are obtained in this phase, and they will be used in the search part of the active shape model.

*Phase II-facial feature extraction*. In this phase, we utilize the ASM (Lai et al., 2001; Christos et al., 2003) to extract the facial chin contour. ASM is composed of two parts: the ASM shape model and the ASM search. However, nonlinear variations caused by the face rotation in the depth direction are not considered in the above models. Due to face rotation in the depth direction, point distribution model (PDM) (Cootes et al., 1995) can only find the exact structure in the new image which has the small rotation. Therefore, some PDMs corresponding to different face poses need to be trained first. Generally, $0^\circ$ corresponds to the frontal face and $-90^\circ$ and $90^\circ$ correspond to the left and right profile face views respectively. The pair of left view and right view is simply mirrors of each other, and therefore only from frontal to the right view is considered and is equally divided into 10 parts. The face samples in the training set are selected to train the PDM about facial chin in each part. In real applications, based on the parameters (of the position, pose, scale) estimated in the face detection phase, the PDM of corresponding pose is chosen. In this way, the exact chin contour is extracted correctly. This is the main idea of the second part.

The rest of this paper is organized as follows. In Section 2, we describe the face detection algorithm and the way face pose is represented. The ASM is discussed in detail in Section 3. The experimental result is reported in Section 4. Conclusions and future directions are discussed in Section 5.

## 2. Related work

The research of face detection and facial feature extraction has been widely used in many applications, but only a few approaches have combined detection with facial feature extraction. In the following, we depict an overview of existing approaches (Liu et al., 2003; Saber and Tekalp, 1998; Shih and Chuang, 2004) in face detection and facial feature extraction as well as put our work in perspective.

### 2.1. Face detection

The role of face detection makes sure the presence and location of a face in an image, by distinguishing the face from all others patterns presented in the scene. But the process of face detection also needs to take into account the following factors:

(1) *Viewing geometry* (*pose*). The images of a face vary due to the relative camera–face pose (frontal, $45^\circ$, profile, upside), and some facial features such as an eye or the nose may become partially or wholly occluded (generally, face exceeds $20^\circ$ in the depth direction rotation).
(2) *Image orientation*. Face images directly vary for different rotations about the camera's optical axis.
(3) *Image conditions*. In the process of the image forming, some factors will affect the quality of image, for example, sensor response, resolution, illumination, focus, imaging nose and man-made factors, etc.
(4) *Facial expression*. Facial expression has an important influence on the appearances of faces. So considering the affect of facial expression is very necessary in the face recognition.
(5) *Presence or absence of structural components*. Facial accessories such as glasses, beards, may or may not be present and there is a great deal of variability among these components including shape, color, and size, etc.

The following reviews the existing techniques that detect faces from a single intensity or color image. Frontal face detection has already relatively perfected by research in the past years, but multi-view face detection still remains a large open problem and has great impact on the multi-view facial feature extraction. We classify single face detection methods into five categories; some methods clearly overlap boundaries.

(1) *Knowledge-based methods*. These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features. In this category, face detection methods are developed based on the rules derived from the research's knowledge of human faces. However, one problem with this approach is the difficulty in translating human knowledge into well-defined rules. The typical example of the approach is the multi-resolution rule-based method (Yang and Huang, 1994).
(2) *Feature invariant approaches*. The goal of these algorithms is to find structural features that exist even when the pose, viewpoint and lighting conditions vary, and then use these features to locate faces in the arbitrary input image. Generally, face features include eyebrows, eyes, nose, mouth, skin color (Hsu et al., 2002) and hair-line extracted by edge detectors. Based on these extracted facial features, a statistical model is built to describe their spatial relationships and to verify the existence of a face. In this category, the problem of

extracting face image features must be taken into account in the application. Specifically these features are usually corrupted due to illumination, image nose etc, and feature boundaries can be weakened for faces when shadows can cause numerous strong edges which together render perceptual grouping algorithms useless. So the image should be preprocessed aiming to facilitate feature extraction.

(3) *Template matching methods*. Several standard patterns of a face are stored in advance to describe the whole face or the separate facial features. The correlations between an input image and the stored patterns are computed for face detection. The advantage of these methods is that they are simple to implement. However, it has been proven that these methods are inadequate for face detection since they cannot effectively deal with variations in scale, pose, and shape. The typical examples have Shape template (Karungaru et al., 2004), etc.

(4) *Appearance-based methods*. In contrast to template matching, the models (or templates) are learned from a set of training images that should capture the representative variability of facial appearance. In general, appearance-based methods mainly depend on techniques from statistical analysis and machine learning to find the relevant characteristics of face and non-face images. The learned characteristics are in the form of distribution models or discriminant functions that are consequently used for face detection. Meanwhile, dimensionality reduction is usually carried out for the sake of computation efficiency and detection efficacy. The typical example of the approaches is the eigenface, distribution-based (Sung and Poggio, 1998) neural network (Rowley et al., 1998), SVM and HMM (Nefian and Hayes III, 1998), etc. But these methods deal with only the type of faces that have been trained or defined on the training set. Viola and Jones (2001) combine a set of efficient classifiers in a cascaded structure to achieve a high speed for frontal face detection, but this approach is not effective for the profile face, due to changing imaging conditions or under occlusion. Most of the literatures in face detection deal with frontal faces. If profile views need to be detected, handing intermediate face pose requires training many different head poses images, thus the training cost will be prohibitive expensive, moreover need a large number of face training samples. In sum, appearance-based methods are more flexible and extensive than other face detection approaches.

(5) *Rapid object detection using a boosted cascade of simple feature*. This method entails a machine learning approach for visual object detection, which is capable of processing images extremely rapidly and achieving high detection rates. First it introduces a new image representation called "integral image" which allows the features used by detector to be computed very quickly. The integral image can be computed from an image using a few operations per pixel. Once computed, any one of these Harr-like features can be computed on any scale or location in constant time. The second is a learning algorithm based on Adaboost, which selects a small number of critical visual features from a large set and yield extremely efficient classifies. The third is a method for combining increasingly more complex classifiers in a "cascade" which allows background regions of the image to be quickly discarded while spends more computation on promising object-like regions.

The object detection classifies are based on the value of simple features. The simple features used are reminiscent of Haar basis functions. Generally, it uses three kinds of features: two-rectangle feature, three-rectangle feature and four-rectangle feature. Rectangle features can be computed very rapidly using an intermediate representation for the image called the integral image. In order to use a small number of features to form an effective classifier, the weak learning algorithm is designed to select the single rectangle feature which best separates the positive and negative examples. For each feature, the weak learner determines the optimal threshold classification function to ensure that the minimum number of examples is misclassified.

The overall form of the detection process is that of a degenerate decision tree, called a "cascade". A positive result from the first classifier triggers the evaluation of a second classifier, which has also been adjusted to achieve a very high detection rate. A positive result from the second classifier triggers the third classifiers, and so on. The cascade training process involves two types of tradeoffs. In most cases, classifiers with more features will achieve higher detection rates and lower false positive rates. At the same time classifiers with more features require more time to compute.

In this paper, we adopt the Adaboost learning method in the process of face detection. The Adaboost algorithm, which was introduced in 1995 by Freund and Schapire (1997) solved many of the practical difficulties of the earlier boosting algorithms. Practically, Adaboost has many advantages in face detection. First, it is fast, simple and easy to implement. It has no parameters to tune and requires no prior knowledge about the weak learner and can be flexibly combined with any method for finding weak classifiers. Secondly, it comes with a set of theoretical guarantees given sufficient data and a weak learner that can reliably provide only moderately accurate weak classifiers. This is a shift in mind set for the learning system designer, instead of trying to design a learning algorithm that is accurate over the entire space, we can instead focus on finding weak learning algorithms that only need to be better than random (Schapire, 2002). Finally, regarding the choice of distribution, the technique that we advocate is to place the largest weight on the examples most often misclassified by the preceding weak rules and this has the effect on forcing the weak learner to focus its attention on the “hardest” examples. As for combining the weak rules, simply taking a (weighted) majority vote of their predictions is natural and effective. The last merit manifests that Adaboost gains an advantage over the boost. In sum, Adaboost based face detection is the most effective approach in face detection approaches.

### 2.2. Face pose estimation

Face pose estimation is a challenging research topic in face detection. Due to face rotation in the depth direction, many difficult problems due to nonlinear changes will occur. In addition, face recognition and tracking are also involved in the face pose problem. Three general approaches have been used to deal with this problem in recent years: (a) To use a full 3D model (Fua and Miccio, 1998; Pighin et al., 1999; Vetter and Blanz, 1998). In these approaches, the 3D face model is firstly built by performing a statistical analysis over a face training set, and the model represents both shape and gray-level appearance. Secondly, fitting the 3D face model to faces in new images uses the robust multi-resolution search algorithm. After fitting, face pose can be settled easily and effectively. The computation cost is very large, which is the main disadvantage of the approach. (b) To introduce non-linearity into a 2D model (Romdhani et al., 1999, 2000). KPCA combining with ASM is typical method to deal with the face pose problem, but the premise of this method is the multi-view face detection, and their approach to constrain shape variability is not generally valid. (c) To use a set of models to represent appearance from different viewpoints (Cootes et al., 2002). If face images have illumination change in the face database, this method cannot gain good result since it is sensitive to illumination.

### 2.3. Facial chin contour extraction

In face recognition, the chin contour is the important facial feature. Due to its weak contrast with neck, chin contour is difficult to extract. But considering its importance, chin extraction should be taken into account.

Facial chin contour extraction approaches differ in the geometry relationship and cues used to model the chin contour. In the geometry relationship-based approach, Vezhnevets et al. (2004) and Arca et al. (2003) use chin shape approximate parabola to extract chin contour. But the premise of the method is the positions and the dimensions of the facial features, and restricts significantly the parabola search area, which is the disadvantage of this method. In addition, invariable graphics cannot build the best model for chin contour, because they cannot adapt to the changes of the illumination, facial pose, and facial expression, this is another reason not to extract facial chin using this method. Therefore, the approach to extract the facial features robustly is difficult. Alternatively chin extraction can be carried out by employing the facial feature spatial geometry relationship. Choudhury Verma et al. (2003) utilize template matching to extract facial features, but this method is not appropriate for the side view face and profile view faces.

### 2.4. Our work in perspective

The following highlights the contribution of our approach and its advantages over the existing approaches. The characteristics of our approach which distinguish it from existing approaches are:

(1) Our approach simultaneously performs face detection and facial pose estimation. Therefore the approach can handle changes in imaging conditions (face scale, lighting, and orientation) and changes in image content (the complexity of the background).
(2) The challenging problem faced with any face detection approaches is the facial pose changes. We adopt a novel pose (Choudhury Verma et al., 2003) representation that can represent any intermediate head pose by combining the information from two detectors, one for frontal views detector, and one for profiles detector.
(3) Most of the chin extraction approaches suffer from the problem of face pose. We develop the active shape model to extract the chin contour.
(4) Through the novel face detection, face angle in the depth rotation is estimated by the combination frontal detection with profile detection, and then chin contour of different angle is extracted easily using the corresponding angle face image training set for ASM.

## 3. Face detection

In this paper, we utilize a wavelet-based probabilistic method to detect face. In the following, we briefly explain the face detector. Furthermore, we introduce a representation for face pose based on a combination of the frontal and the profile detectors.

### 3.1. Detection algorithm

We describe a statistical method for 3D object detection and represent the statistics of both object and “non-object” appearance using a product of histograms. Statistics are based on joint probabilities of visual attributes, and they can provide a facial representation that is jointly localized in space, frequency and orientation. These attributes are obtained by combining quantized wavelet coefficients at different positions and in different frequency bands. To do so, we perform a wavelet transform of the face image. The merit of wavelet transform is that it produces no redundant information. Unlike other transforms, we can perfectly reconstruct the image from its transform where the number of transform coefficients is equal to the original number of image pixels. The wavelet transform organizes the image into sub-bands that are localized in different orientations and different frequencies. We use a wavelet transform based on 3 level decomposition using a 5/3 linear phase filter-bank to produce 10 sub-bands. Each level in the transform represents a higher octave of frequencies. Within each sub-band, each coefficient is spatially localized. We remove the lower frequencies, thus removing light and skin color information. We use different combinations of eight coefficients to form 11 different visual attributes. They are computed at different positions, which are uniformly distributed in the region of the face. For more details, the readers can refer to Schneiderman and Kanade (2000).

The probability of detection is computed as a weighted combination of the probabilities of the 11 visual attributes, evaluated on 16 regions of the image for the face and the non-face distribution. The detector finds a face at a given location and a given scale. In order to detect frontal and profile views, we utilize two detectors are $P_f(Face|I, x, y, s)$ and $P_p(Face|I, x, y, s)$, respectively. The coordinates $x$, $y$ represent the position on which the detector is applied and $s$ is the scale of detection. Implementation details can be found in Choudhury Verma et al. (2003).

The main drawback of a histogram is that we can only use a relatively small number of discrete values to describe object appearance. To overcome this limitation, we use multiple histograms where each histogram $P_k(pattern|object)$ represents the probability of object appearance over some specified visual attribute $pattern_k$; that is a random variable describing some chosen visual characteristic such as low frequency content. We will specify how we partition appearance into different visual attribute. To combine probabilities to form different attributes, we will take the following product where we approximate each class-conditional probability function as a product of histograms:

$$P(\text{image}|\text{object}) \approx \prod_k P_k(pattern_k|object), \tag{1}$$

$$P(\text{image}|\text{non-object}) \approx \prod_k P_k(pattern_k|non\text{-}object). \tag{2}$$

We compute these statistical values from various sets of images. This process of gathering statistics is usually referred to as training. In this paper, we use Adaboost algorithm mentioned above to train classification. Adaboost works in an iterative fashion. First, we train a detector by assigning the same weight to all training examples. Then we iteratively retrain the detector at each iteration where more weight is given to training examples that were incorrectly classified by the detector trained in the previous iteration. Through the iterative process, the classification error can be decreased (Freund and Schapire, 1997; Shapire and Singer, 1999). The whole classifications are trained successfully.

### 3.2. Pose representation

Pose changes are the most important source of variation in faces. The detection of faces from any viewpoint can be solved by learning a multi-view face model successfully. In order to avoid the tedious work to train all possible views and detect faces from any viewpoint, we have combined two detectors, a frontal one and a profile one. This pose estimation is sufficient for predicting its degree in depth. In the process of face detection, we can combine the probabilities $P_P$ and $P_f$ to approximately estimate face pose:

$$\theta = \max\left(\frac{1-p_f}{p_f+p_p}, \frac{p_p}{p_f+p_p}\right) \times 90. \tag{3}$$

The following utilize UMIST face database to estimate the facial pose using the above approach. We will introduce the face database in detail in the experiment section. This face database is publicly available. Fig. 1 shows the parts image of the UMIST face database.

It can be seen that the approach is feasible through the above experiment, and the high calculation cost is the only disadvantage of the approach. Fig. 2 represents the facial

Fig. 1. The face image of the UMIST face database.

| Frame | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_f$ | 0 | 0.05 | 0.19 | 0.28 | 0.34 | 0.43 | 0.51 | 0.59 | 0.64 | 0.69 |
| $P_p$ | 1 | 0.95 | 0.76 | 0.69 | 0.64 | 0.48 | 0.40 | 0.33 | 0.29 | 0.22 |
| $\theta$ | 90 | 85.5 | 76.7 | 66.8 | 60 | 56.37 | 48.46 | 40.11 | 34.83 | 30.65 |
| **Frame** | **10** | **11** | **12** | **13** | | | | | | |
| $P_f$ | 0.77 | 0.85 | 0.91 | 1 | | | | | | |
| $P_p$ | 0.14 | 0.12 | 0.09 | 0 | | | | | | |
| $\theta$ | 22.7 | 13.9 | 8.10 | 0 | | | | | | |

Fig. 2. The detection probability of frontal face and profile face using $P_f$, $P_p$ representation, respectively, and the face rotation angle $\theta$ using Eq. (3).

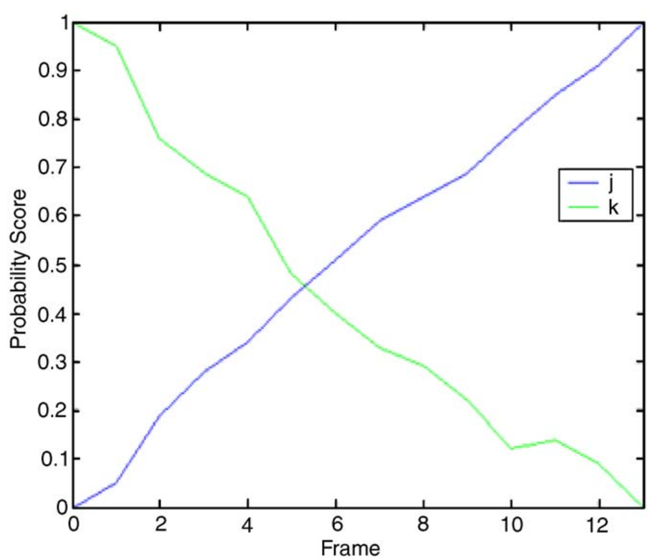


Fig. 3. The frontal and profile face detection probability represented with blue curve and green curve, respectively.

pose estimation result of Fig. 1 image using the proposed method. We also see that the probability of the profile view detector decreases as the face turns toward the camera and the probability of the frontal face detector increases gradually. Fig. 3 shows the change of two detector probability using the graph.

## 4. Active shape model (ASM)

This section briefly reviews the ASM, and then we describe the algorithm of facial chin contour extraction. An ASM consists of a PDM aiming to learn the variations of valid shapes and a statistical model of the grey-levels expected around each landmark.

### *4.1. Point distribution model*

An object is described by $n$ points, referred to as landmark points. The localization of the landmarks can be obtained either by an automatic system, such as a facial feature tracking system developed by Maurer and von der Malsburg (1996), or by the user with the aid of a user-interface system. Fig. 4 shows three examples of labeled face image, where $n$ landmark points are evenly selected from the chin contour of face to constrain the representation of shape model. From the landmark points, a PDM is constructed as follows. The landmark points $(x_1, y_1)$, $(x_2, y_2) \ldots (x_n, y_n)$ are stacked in shape vectors:

$$X = (x_1, y_1, x_2, y_2, \ldots, x_n, y_n).$$

Firstly, the training shapes are aligned to a common coordinates by a transformation that include translation, rotation and scaling. The parameters of the transformation are obtained by minimizing a weighted sum of squares of distances between equivalent points on different shapes. This is a modification of the Procrustes method (Gower, 1975).

Secondly, the set of training shape is aligned, and we can calculate the mean $\bar{x}$ of the aligned training shape set (Kim and Lee, 2003).

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{4}$$

Then deviations from the mean are calculated for each shape of the training shape set.

$$\mathrm{d}x_i = x_i - \bar{x}. \tag{5}$$

We can calculate the $2n \times 2n$ covariance matrix $S$, using

$$S = \frac{1}{N}\sum_{i=1}^{N} \mathrm{d}x_i\, \mathrm{d}x_i^{\mathrm{T}}. \tag{6}$$

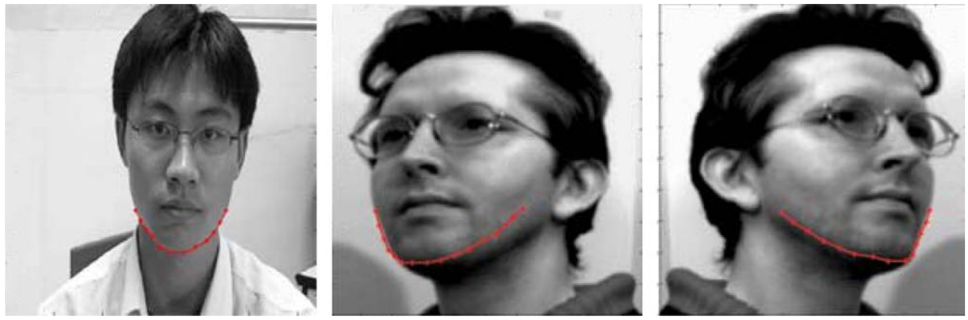

Fig. 4. Training set of face, each chin contour defined by $n$ points.

We perform principal component analysis on $S$ to reduce dimensionality. A small number of modes $t$ can replace the whole variation. Eigenvectors is defined by corresponding eigenvalue.

So a shape model can be represented by

$$x = \bar{x} + Pb, \tag{7}$$

where $\bar{x}$ is the mean shape of the aligned training set, $b = (b_1, b_2, \ldots, b_t)^{\mathrm{T}}$ is a vector of shape parameters, and $P = (P_1, P_2, \ldots, P_t) \in R^{2n\times t}$ is the set of the eigenvectors corresponding to the largest $t$ eigenvalues of the covariance matrix $S$ of the training shape.

According to Eq. (7), new examples of the shapes are generated by varying the parameters ($b_k$) within suitable limits. By experiments, we find that most of the population lies within three standard deviations of the mean, so $|b_i| \leqslant 3\sqrt{\lambda_i}$ is a suitable limit. A new shape will be similar to those in the training set.

### 4.2. Using PDM to locate instance of such shapes in a new image

Since the flexible models have been generated in the above section, the following is to locate instance of such shapes in a new image using the flexible models. Shape and pose parameters will be involved in the search process, and the change of parameters will be adjusted to coincide with the structure of interest in the image. By choosing a set of shape parameters $b$ for a PDM, we define the shape of model object in an object centered coordinate frame. We can create an instance $X$ of the model in the image frame by defining the position, orientation and scale:

$$X = M(s, \theta)[x] + X_{\text{cen}}, \tag{8}$$

where $M(s, \theta)[\,]$ is a rotation by $\theta$ and a scaling by $s$. $X_{\text{cen}} = (X_c, Y_c, X_c, Y_c, \ldots, X_c, Y_c)^{\mathrm{T}}$, $(X_c, Y_c)$ is the center position of the model in the image frame, $x$ denotes in the shape space, and $X$ denotes in the image space.

In this section, we present an iterative approach to find the appropriate $X$ that gives a very rough starting approximation. The starting value of $X$ does not need to be very close to the final solution, therefore the presented approach can be used in many applications.

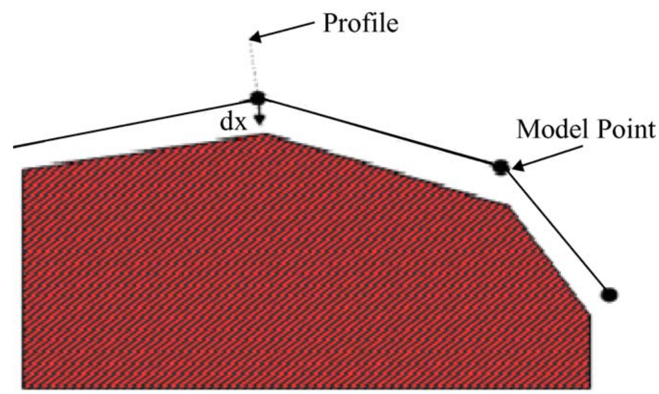


Fig. 5. Suggested movement of point is along normal to boundary, proportional to maximum edge strength on normal.

The following describes an iterative method of local optimization, which fits a model instance to a structure in an image. The iterative process is described as follows:

(1) The current estimation of $X$ is placed into the image and examined a region of the image around each model point to find the best near neighbour location for this point. This must adjust the shape parameters to transform the PDM shape model.
(2) Update the shape model parameters to move the model points toward these new optimal points. These model parameters have some limitations, because global shape constraints can be applied ensuring the shape of the model example remains similar to those of the training set.

The procedure is repeated until convergence. Because the models attempt to better fit the data, it deforms in the way that are consistent with the shapes in the training set. Fig. 5 shows the adjustment along a normal to the model boundary toward the strongest image edge, with magnitude proportional to the strength of the edge.

#### 4.2.1. Calculating the adjustments to the pose and shape parameters

The motive to adjust the pose and shape parameters moves the points along the profile from their current locations in the image frame $X$ to be as close to the suggested new locations ($X$+d$X$). Simultaneously, the model must satisfy the constraint that the shape of model holds the similarity. The current estimate of the model is centered at $(X_c, Y_c)$ with orientation $\theta$ and scale $s$, and we need to update these parameters to better fit the image. The required pose adjustment is achieved by finding the translation $(\mathrm{d}X_c,\ \mathrm{d}Y_c)$, rotation $\mathrm{d}\theta$ and scaling factor $(1+\mathrm{d}s)$ which best maps the current set of points $X$ onto the set of points given by $(X+\mathrm{d}X)$.

The following discuss how to calculate d$X$. The initial position of the points in the image frame is given by Eq. (9)

$$X = M(s,\theta)[x] + X_c. \tag{9}$$

We wish to calculate a set of residual adjustments d$x$ in the local model coordinate frame such that

$$M(s(1+\mathrm{d}s),(\theta+\mathrm{d}\theta))[x+\mathrm{d}x] + (X_c + \mathrm{d}X_c) = (X + \mathrm{d}X). \tag{10}$$

Thus

$$M(s(1+\mathrm{d}s),(\theta+\mathrm{d}\theta))[x+\mathrm{d}x] = (M(s,\theta)[x] + \mathrm{d}X) - (X_c + \mathrm{d}X_c), \tag{11}$$

and since

$$M^{-1}(s,\theta)[\,] = M(s^{-1},-\theta)[\,], \tag{12}$$

we obtain

$$\mathrm{d}x = M((s(1+\mathrm{d}s))^{-1}, -(\theta+\mathrm{d}\theta))[y] - x, \tag{13}$$

where

$$y = M(s,\theta)[x] + \mathrm{d}x - \mathrm{d}X_c.$$

Eq. (13) calculates the suggesting movements along the profile to the point $x$ in the local model coordinate frame. In order to utilize the shape constraints, we transform d$x$ into the model parameter space. The adjustments to the shape parameters d$b$, which will best match the model to the suggested new positions given by solving. So we search the relationship of d$x$ and d$b$.

Eq. (7) gives

$$x = \bar{x} + Pb.$$

We wish to find db such that

$$x + \mathrm{d}x = \bar{x} + P(b + \mathrm{d}b). \tag{14}$$

Since there are only $t(t<2n)$ modes of variation available and d$x$ can move the points in $2n$ different degrees of freedom, we can only achieve an approximation to the deformation required.

Subtracting Eq. (7) from Eq. (14) gives

$$\mathrm{d}x = P(\mathrm{d}b).$$

Therefore

$$\mathrm{d}b = P^{\mathrm{T}}\,\mathrm{d}x. \tag{15}$$

Since the columns of $P$ are mutually orthogonal and of unit length we know that $P^{\mathrm{T}} = P^{-1}$. It can be shown that Eq. (15) allows us to calculate changes to the shape parameters d$b$ required to improve the match between an object model and image evidence. When these changes are applied we can ensure that the model only deforms into shapes consistent with the training set by placing limits on the values of d$b$. A new example can be calculated and new suggested movements can be derived for each point.

### 4.2.2. *Updating the pose and shape parameters*

The equations mentioned above allow us to calculate changes to the pose variables and adjustments, $\mathrm{d}X_c$, $\mathrm{d}Y_c$, $\mathrm{d}\theta$, d$s$ and d$b$. We apply these to update the parameters in an iterative scheme,

$$\begin{aligned} X_c &\to X_c + w_t\,\mathrm{d}X_c \\ Y_c &\to Y_c + w_t\,\mathrm{d}Y_c \\ \theta &\to \theta + w_\theta\,\mathrm{d}\theta \\ s &\to s(1 + w_s\,\mathrm{d}s) \\ b &\to b + w_b\,\mathrm{d}b, \end{aligned} \tag{16}$$

where $w_t$, $w_s$, and $w_\theta$ are scalar weights, and $w_b$ is a diagonal matrix of weights. In this paper, $w_t$, $w_s$ and $w_\theta$ initializations are set 1, $w_b$ is initialized with the unit matrix, and $X_c$, $Y_c$, $\theta$, $s$ is provided by the above face detection.

### 4.2.3. *The strategy of search*

In search method (Cootes et al., 1994), we use multi-resolution search method. Generally, the computation time is affected by two factors: search data size and search space. Multi-resolution image generally applies a coarse-to-fine resolution approach to reduce the size of the search data. Using the coarse-to-fine grain scheme, multi-resolution techniques reduce the size of the search data by searching initially at lowest resolution (smallest size) first and then proceeding to higher resolution where the search results are only refined. We can generate such images from the original images using Gaussian smoothing and sub-sampling to produce a multi-resolution pyramid. Level 0 in such a pyramid is the origin image, and level 1 is an image with half the number of pixels along each axis. In our approach, each pixel in a given level is generated by smoothing the image at the level below with a $5\times5$ Gaussian mask.

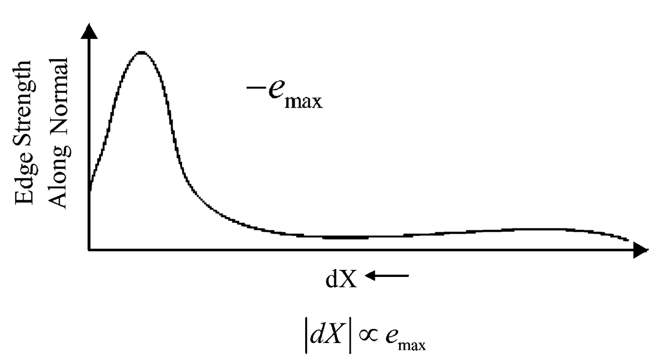


Fig. 6. Optimal match position of the movement along profile.

In this paper, we regularized the coordinate size of landmarks along with different resolution. According to the initial estimation of the positions of a set of landmarks that we attempt to fit to an image object, we need to find a set of adjustments that will move each point toward a better position. We use an adjustment along profile toward the strongest edge points (Fig. 6).

## 5. Experimental results

In this section, we present the experiment results carried out on different face database with single face occurring in different poses. In our study, we use three face databases to test our proposed approach. The first one is the UMIST database, the second one is the SJTU database, and the last one is the FERET database.

### 5.1. Face database

The first face database is publicly available. The UMIST Face Database (http://images.ee.umist.ac.uk/danny/database.html) consists of 546 grayscale images of 20 people of both genders and various races (image size is about $220 \times 220$). Various pose angles of each person are not provided accurately, ranging from profile to frontal views. The images of each person are numbered consecutively, in the order when they were taken. While this provides some indication of relative pose angle in each image, this is not always reliable perhaps due to the movement of the subject between captures. So no absolute pose angle is provided for each image.

Considering domestic practical need, we use SJTU face database as the second face database. SJTU face database contains 202 female and male people. Each subject holds nine different facial pose, ranging from the left profile to the right profile, and each face image has not accurate pose angle. The whole face database consists of 1818 gray images.

In order to further verify the effective of the proposed approach, we utilize the third face database FERET face database (http://www.itl.nist.gov/iad/humanid/feret/feret_master.html). A set of images collected by the US Army Research Laboratory to develop, test, and evaluate face recognition algorithms and face pose estimation algorithms, etc. The database consists of greyscale (with 8 bits per pixel) and color images with the size of $512 \times 768$, and the face rotation angle in the depth is offered accurately. In our approach, we only use the grayscale image for the experiment.

### 5.2. Experimental results

In UMIST face database, from frontal face to profile face, facial rotation angle in the depth is $90^\circ$. In this experiment, we only consider from frontal to the right view, and then $90^\circ$ is then equally divided into 10 parts. In each part, facial chin contours of 100 training object are first manually labeled, and these objects are trained to obtain the chin flexible model (PDM). So 10 PDM are obtained in the training, which is the case of face rotating from frontal to right profile. Due to the symmetry, we can obtain the part of face rotating from frontal to left profile through mirroring projection. Having generated point distribution models, we would like to use them in image search to find new examples of modeled objects in images. We utilize the multi-resolution method to search the structure of interesting in the image coinciding with the PDM. We select randomly 100 face images as testing samples. The following is part of experimental results on the UMIST face database ( Fig. 7).

Considering domestic application, our SJTU face database is used. Due to race difference, face configuration have difference in many aspects. Each face has nine different facial poses in the database. We select 100 samples in different facial poses as the training set and then select randomly 100 face samples as the testing set.

It can be seen from Fig. 8 that the face pose estimation approach is effective, and the facial chin can be extracted accurately based on the face pose estimation.

In order to further evaluate our method, experiments are conducted on the FERET face database. We select 50 face samples with the same pose as training set, and select randomly 100 face samples as test samples. The part of experimental results is shown in Fig. 9. It can be seen from Table 1 that there are the comparison results about chin contour extraction in three face database.

Through experimental results on the three face database above, we can see that the method is valid. Due to space

Fig. 7. Facial chin extraction in UMIST face database.

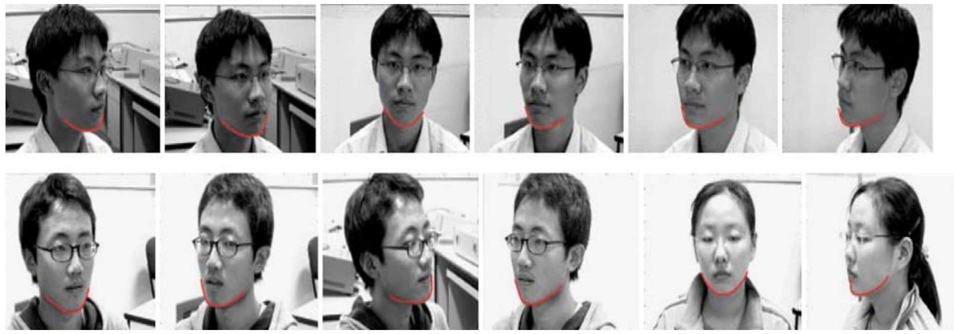

Fig. 8. Facial chin extraction in SJTU face database.

Fig. 9. Facial chin extraction in FERET face database.

Table 1
Comparison of experimental result in three different face database

| Face database | Training set (certain within pose range) | Testing set | Chin extraction accuracy (%) | Average accuracy |
|---|---|---|---|---|
| UMIST | 100 | 100 | 95 | 95.3% |
| SJTU | 100 | 100 | 95 | |
| FERET | 100 | 100 | 96 | |

limitation, we use the following table to show the all experimental results.

## 6. Conclusion and in the future

In this paper, we firstly present the multi-view face detection based on a combination of the frontal and the profile detectors and describe a new representation for pose. We can accurately detect the face in the face image, using new pose representation method to estimate the face pose.

Secondly, we introduce active shape model (ASM). An ASM consists of a point distribution model (PDM) aiming to learn the variations of valid shapes, and a set of flexible models capturing the gray-levels around a set of landmark feature points. We train together 10 flexible models according to different face rotation angle.

In the process of extracting the facial chin contour in face image. Firstly, we detect face in the image, using the pose representation to estimate the facial pose. We know the angle of facial rotation in depth direction, and corresponding PDM is found. We can search the interesting structure in the image.

In the future, we use facial feature points to establish the point correspondence from each face image in the training set and a reference image, which is the core of the building morphable model. On the basis of the morphable model, we perform the 3D face recognition. Face recognition system performance will be more increased greatly than existing approaches.

## Acknowledgement

The research is sponsored by Fundamental Project of Committee of Science and technology, Shanghai under contracts 03DZ14015. The authors would like to thank Dr. Z. L. Zheng for useful discussion.

## References

Arca, S., Campadelli, P., Lanzarotti, R., 2003. A face recognition system based on local feature analysis. Fourth International Conference on AUDIO- and VIDEO-BASED BIOMETRIC PERSON AUTHENTICATION, June.

Choudhury Verma, R., Schmid, C., Mikolajczyk, K., 2003. Face detection and tracking in a video by propagating detection probabilities. Pattern Analysis and Machine Intelligence 25 (10), 1215–1228.

Christos, D., Xiaodong, T., Dinggang, S., 2003. Hierarchical active shape models, using the wavelet transform. IEEE Transactions on Medical Image 22 (3), 414–423.

Cootes, T.F., Taylor, C.J., Lanitis, A., 1994. Multi-resolution search with active shape models. Pattern Recognition (1), 610–612.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—their training and application. Computer Vision and Image Understanding 61 (1), 38–59.

Cootes, T.F., Wheeler, G.V., Walker, K.N., Taylor, C.J., 2002. View-based active appearance models. Image and Vision Computing 20 (8), 657–664.

Feraud, R., Bernier, O.J., Viallet, J.-E., Collobert, M., 2001. A fast and accurate face detector based on neural networks. IEEE Transactions Pattern Analysis and Machine Intelligence 23 (1), 42–53.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55 (1), 119–139.

Fua, P., Miccio, C., 1998. From regular images to animated heads: a least squares approach. In: Burkhardt, H., Neumann, B. (Eds.), Fifth European Conference on Computer Vision 1. Springer, Berlin, pp. 188–202.

Gower, J.C., 1975. Generalized Procrustes analysis. Psychometrika 40, 33–51.

Hsu, R.-L., Mohamed, A.-M., Jain, A.K., 2002. Face detection in color images. Pattern Analysis and Machine Intelligence 24 (5), 696–706.

Karungaru, S., Fukumi , M., Akamatsu, N., 2004. Feature extraction for face detection and recognition. Proceedings of the IEEE Conference on Robot and Human Interactive Communication, pp. 235–239.

Kim, W., Lee, J.-J., 2003. Shape tracking based on the modular active shape model. Advanced Intelligent Mechatronics 6 (2), 1411–1416.

Koh, L.H., Ranganath, S., Lee, M.W., Venkatesh, Y.V., 1999. An integrated face detection and recognition system image analysis and processing, 1999. Proceedings International Conference vol. 9, 532–537.

Kumar, N., Abhishek, V., Gautam, G., 2003. A novel approach for person authentication and content-based tracking in videos using kernel methods and active appearance models. Systems, Man and Cybernetics, 2003. IEEE International Conference 2 (10), 1384–1389.

Lai, J.H., Yuen, P.C., Feng, G.C., 2001. Face recognition using holistic Fourier invariant features. Pattern Recognition 34 (1), 95–109.

Lie, G., Li, S.Z., Zhang, H.-J., 2001. Learning probabilistic distribution model for multi-view face detection. Computer vision and pattern recognition. Proceedings of the 2001 IEEE Computer Society Conference 2 (12), II-116–II-122.

Liu, C., 2004. Gabor-based kernel PCA with fractional power polynomial models for face recognition. Pattern Analysis and Machine Intelligence 26 (5), 572–581.

Liu Z.-F., You, Z.-S., Jain, A.K., Wang, Y.-Q., 2003. Face detection and facial feature extraction in color image, Computational intelligence and multimedia applications, 2003. ICCIMA 2003. Proceedings of the Fifth International Conference on 27–30 September, pp. 126–130.

Maurer, T., von der Malsburg, C., 1996. Tracking and learning graphs and pose on image sequences. In: Proceeding of the International Workshop on Automatic Face and Gesture Recognition. Vermont, pp. 176–181.

Nefian, A.V., Hayes III, M.H., 1998. Face detection and recognition using hidden Markov models. Proceedings of the IEEE International Conference on Image Processing, pp. 141–145.

Othman, H., Aboulnasr, T., 2003. A separable low complexity 2D HMM with application to face recognition. Pattern Analysis and Machine Intelligence 25 (6), 1229–1238.

Perlibakas, V., 2004. Distance measures for PCA-based face recognition. Pattern Recognition Letter 25 (6), 711–724.

Pighin, F., Szeliski, R., Salesin, D.H., 1999. Resynthesizing facial animation through 3d model-based tracking. Seventh International Conference on Computer Vision, pp. 143–150.

Romdhani, S., Gong, S., Psarrou, A., 1999. A multi-view non-linear active shape model using kernel pca. In: Pridmore, T., Elliman, D. (Eds.), 10th British Machine Vision Conference 2. BMVA Press, Nottingham, pp. 483–492.

Romdhani, S., Psarrou, A., Gong, S., 2000. On utilising template and feature-based correspondence in multi-view appearance models. Sixth European Conference on Computer Vision 1. Springer, Berlin, pp., 799–813.

Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural network-based face detection. Pattern Analysis and Machine Intelligence 20 (1), 23–38.

Saber, E., Tekalp, A.M., 1998. Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. Pattern Recognition Letter 19 (8), 669–680.

Schapire, R.E., 2002. The boosting approach to machine learning: an overview. In: MSRI Workshop on Nonlinear Estimation and Classification.

Schneiderman, H., Kanade, T., 2000. A statistical method for 3D object detection applied to faces and cars. Proceedings of the Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 746–751.

Shapire, R.E., Singer, Y., 1999. Improving boosting algorithms using confidence-rated predictions. Machine Learning 37 (3), 297–336.

Shih, F.Y., Chuang, C.-F., 2004. Automatic extraction of head and face boundaries and facial features. Information Sciences 158 (1), 117–130.

Sung, K.-K., Poggio, T., 1998. Example-based learning for view-based human face detection. Pattern Analysis and Machine Intelligence 20 (1), 39–51.

Vetter, T., Blanz, V., 1998. Estimating coloured 3d face models from single images: an example based approach. In: Burkhardt, H., Neumann, B. (Eds.), Fifth European Conference on Computer Vision 2. Springer, Berlin, pp. 499–513.

Vezhnevets, V., Soldatov, S., Degtiareva, A., Park, I.-K., 2004. Automatic extraction of frontal facial features. Proceedings of the Sixth Asian Conference on Computer Vision (ACCV04) 2, Jeju, Korea, January, pp. 1020–1025.

Viola, P., Jones, M.J., 2001. Robust real-time object detection. In the Proceedings of IEEE Workshop on Statistical and Theories of Computer Vision.

Volker, B., Thomas, V., 2003. Face recognition based on fitting a 3D morphable model. Pattern Analysis and Machine Intelligence 25 (9), 1063–1074.

Yang, G., Huang, T.S., 1994. Human face detection in complex background. Pattern Recognition 27 (1), 53–63.

Yang, M.-H., Kriegman, D.J., Ahuja, N., 2002. Detecting faces in images: a survey. IEEE Transactions Pattern Analysis and Machine Intelligence 24 (1), 34–58.

Yongmin, L., Shaogang, G., Sherrah, J., Liddell, H., 2000. Multi-view face detection using support vector machines and eigenspace modeling. Knowledge-based intelligent engineering systems and allied technologies, 2000. Proceedings Fourth International Conference 1 (9), 241–244.

Yuille, A.L., Hallinan, P.W., Cohen, D.S., 1992. Feature extraction from faces using deformable templates. International Journal of Computer Vision 8, 99–111.