

Improved Active Shape Model for Facial Feature Extraction in Color Images

Mohammad H. Mahoor, Mohamed Abdel-Mottaleb[†], and A-Nasser Ansari
Department of Electrical and Computer Engineering, University of Miami

1251 Memorial Drive, Coral Gables, FL 33146

Emails: mmahoor@umsis.miami.edu, mottaleb@miami.edu, and a.alansari@umiami.edu

[†]Corresponding author

Abstract—In this paper we present an improved Active Shape Model (ASM) for facial features extraction. The original ASM developed by Cootes *et al.* [1] suffers from factors such as, poor model initialization, modeling the intensity of the local structure of the facial features, and alignment of the shape model to a new instant of the object in a given image using simple Euclidian transformation. The core of our enhancement relies on three improvements (a) initializing the ASM model using the centers of the mouth and eyes, which are located using color information, (b) incorporating RGB color information to represent the local structure of the feature points, and (c) applying 2D affine transformation in aligning the facial features that are perturbed by head pose variations, which effectively aligns the matched facial features to the shape model and compensates for the effect of the head pose variations. Experiments on a face database of 70 subjects show that our approach outperforms the standard ASM and is successful in extracting facial features.

Index Terms—shape model, color image, face recognition, facial feature extraction, lips detection.

I. INTRODUCTION

Face recognition has become one of the most important applications of image analysis and computer vision in recent years. This trend caught the attention of many academic and industrial research groups [2]. In [2], the methods for face recognition are divided into two main categories: holistic approaches, and geometric-based approaches. In the holistic methods, the recognition is achieved by modeling or representing the intensity values of the pixels in facial images, while in the geometric methods, the geometry of the face and the locations of salient points, known as facial features, are taken into account. In the following subsection we review the literature on facial feature extraction.

A. Literature Review

Facial feature extraction for face recognition is a challenging problem. There are many methods for facial feature extraction [3]–[10]. Table I reviews some of the important works published for facial feature extraction. Kobayashi *et al.* in [3] described an automated algorithm for face detection and facial feature extraction from video images with free backgrounds. The extracted features are points around the eyes, mouth, nose and facial contours.

The authors used spatiotemporal difference images to extract these feature points. For example blinking is used for detecting the eyes. The method proposed by De Natalie *et al.* in [4] aimed at identifying the position of characteristic facial elements (eyes, nose and mouth). The proposed detection strategy is based on the identification of the face symmetry axis, and the successive detection of eyes, mouth and other relevant facial features using correlation. Nikolaidis *et al.* in [5] described a method for extracting facial features with the goal of 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 gray level pixels in the image. Candidates for cheeks and chin are determined by performing adaptive hough transform on a relevant sub-image defined according to the position of the eyes, mouth, and the ellipse containing the main connected component of the image. In order to acquire a more accurate model of the face, a deforming technique is also applied to the ellipse representing the main face region. Candidates for eyebrows are determined by adapting a proper gray level template to an area restricted by the position of the eyes.

Lam *et al.* [11] devised an efficient approach for detecting and locating the eyes in frontal images. Possible eye candidates in an image are identified by means of the valley features and corners of the eyes. Two possible eye candidates are considered to belong to the eyes of a human face if their respective local properties are similar; an eye window is then formed. Each of the eye region candidates is then further verified by comparing them with a standard eye template, and by measuring its symmetry. Yagi [12] presented a system that integrates a library of 32 functions for automatic facial contour extraction. The 32 functions can be classified into five groups such as face detection, pupil detection, facial parts detection, facial parts contour extraction, and face contour extraction. The system is not only useful for the automatic 3D facial model fitting, but also for a range facial image processing applications such as personal authentication and facial expression analysis. Lau *et al.* in [6] proposed an energy function which is the sum of seven weighted terms for

TABLE I.
FACIAL FEATURE EXTRACTION METHODS.

Author	Face detection	Approach	Number of Feature Points	Video/Still
Seiji Kobayashi et al 1995 [3]	Included	Using spatiotemporal difference image to extract the feature points	Representing eyes, eyebrows, boundary, mouth and face boundary by contours	Video - Frontal
T.F. Cootes et al 1996 [1,20]	-	Using active shape models	Flexible in term of number of features	Still - Frontal
Francesco G.B. De Natalie 1997 [4]	Included	Identification of the face symmetry axis and then feature points using correlation.	Location of eyes, and mouth	Still - Frontal
Alhanasios Nikoladis 1997 [5]	Included	Using adaptive Hough transform, template matching and active contour models	Eyes, eyebrows, mouth, nostrils, cheeks and chin	Still - Frontal
Kin-Man Lam et al 1998 [11]	-	Eye corner detection and template matching	Eyes	Still - frontal
Yashushi Yagi 2000 [12]	Included	-	Pupil, and contours	Still - Frontal
C.M. Lau et al. 2001 [6]	Included	Using energy function to extract facial features	Face region, iris, eyebrows, nose, mouth, eye corners, face-hair	Still - Frontal
Gary G. Yen et al. 2002 [7]	Included	Using the edge density distribution of the image and genetic algorithm	Face region, eyes, nose, mouth	Still - Frontal
Kap-Ho Seo et al. 2002 [8]	Included	Using active contour model and color information	Face region, eyes, nose, mouth	Still - Frontal
Dihua Xi et al. 2002[9]	Included	Using Support Vector Machines and Multi wavelet decomposition	Face region, eyes, nose, mouth	Still - Frontal
Zhong Xue et al 2002 [13]	-	Using Bayesian shape model	A mesh model of facial features	Still - Frontal
Rein-Lien Hsu et al. 2002 [14]	Included	Using color information for face detection	Face boundary, eyes, mouth	Still - Frontal with pose
Yongli Hu et al 2003 [15]	-	Using linear combination model	Eyes, eyebrows, nose, mouth	Still - Frontal
Aysegul Gundaz et al 2003 [16]	-	Using topological operators	Eye corners and mouth center	Still - Frontal
Kyung-A Kim et al 2004 [10]	-	Using PCA and Wavelet multi resolution images	Eyes, eyebrows, nose, mouth	Still - Frontal
Kenji Nagao 2004 [17]	-	Using Bayesian approach with nonlinear kernels	Eye centers	Still - Frontal with pose

facial feature extraction. By allocating different values for the weighting factors, the function can extract different fiducial points.

Yen *et al.* in [7] presented a method for facial feature extraction that uses the edge density distribution of the image. In the preprocessing stage a face is approximated to an ellipse, and a genetic algorithm is applied to search for the best matching region. In the feature extraction stage, a genetic algorithm is applied to extract the facial features, such as the eyes, nose and mouth, in the pre-defined sub regions. The authors validated their method by experimenting on various video images under natural lighting environments and in the presence of noise and different face orientations. Seo *et al.* in [8] presented an active contour model based upon color information for extracting facial features. Their algorithm is composed of three main parts: the face region estimation part, the detection part and the facial feature extraction part. In the face region estimation part, images are segmented based on human skin color. In the face detection part, a template matching method is used, and in the facial feature extraction part, an algorithm called “color

snake” is applied to extract facial feature points within the estimated face region. Xi *et al.* in [9] developed an algorithm for detecting human face and extracting facial features. For this task, a flexible coordinate system and several support vector machines were developed for both face detection and extraction based on multi-resolution wavelet decomposition (MWD). Xue *et al.* in [13] presented a novel application of the Bayesian Shape Model (BSM) for facial feature extraction. First, a full-face model is designed to describe the shape of a face, and the PCA is used to estimate the shape variance of the face model. Then, the BSM is applied to match and extract the face patch from input face images. Finally, using the face model, the extracted face patches are easily warped or normalized to a standard view. Hsu *et al.* in [14] proposed a face detection algorithm from color images in the presence of varying lighting conditions as well as complex backgrounds. Based on a novel lighting compensation technique and a nonlinear color transformation, this method detects skin regions over the entire image and then generates face candidates based on the spatial arrangement of these skin patches. The

algorithm constructs eye, mouth, and boundary maps for verifying each face candidate. Hu *et al.* in [15] proposed a facial feature extraction method based on a linear combination model. The model uses the knowledge of prototype faces, which are manually labeled, to interpret novel faces. Generally, the construction of the linear combination model depends on pixel-wise alignments of the prototypes, and the alignments are computed by an optical flow algorithm or bootstrapping algorithm which is a full-scale optimization and without including local information such as facial feature points.

Gundaz *et al.* in [16] presented a method for facial feature extraction by considering the face image as a surface. Topological properties of the facial surface, such as principal curvatures are used to extract the eyes and mouth, which form deep valleys on the surface. The basic idea of the proposed method is to model the facial features as ravines on the facial surface. Ravines are points on the surface where the maximum curvature is a local maximum in the corresponding principal direction. Kim *et al.* in [10] proposed an algorithm for extraction of the facial feature (eyebrow, eye, nose, and mouth) fields from gray level face images. The foundation of this method is that eigenfeatures, derived from the eigenvalues and eigenvectors of the gray-level data set constructed from the feature fields, are very useful to locate these fields efficiently. In addition, multi-resolution images, derived from a 2-D DWT (Discrete Wavelet Transform), are used to reduce the search time for the facial features. Nagao in [17] described a method for finding the positions of features in facial images. A large class of image variations, including those resulting from object rotation in 3D space and scaling (i.e. translation in depth) are handled. A MAP (Maximum a Posteriori) estimation technique using Gaussian distribution is exploited to model the relationship between images and feature positions.

In the area of non-rigid object segmentation the so called Deformable Models have received much attention in the recent years. These models have proven to be efficient in many applications such as object segmentation, appearance interpretation, motion tracking etc. A deformable model can be characterized as a model, which under an implicit or explicit optimization criterion deforms the shape to match a known object in a given image. For a general review of the most commonly used models refer to [18], [19]. Cootes *et al.* [1], [20] developed a statistical approach called Active Shape Model (ASM) for shape modeling and feature extraction. In the following subsection, we review the ASM method and address its problems.

B. Active Shape Model and its Limitations

Active Shape Model (ASM) is a statistical approach for shape modeling and feature extraction. It represents a target structure by a parameterized statistical shape model obtained from training. This method was introduced by Cootes *et al.* [1], [20] and improved by other researchers over the past few years. In the original version of the

ASM, the initial feature points are obtained from the mean shape, which is derived from training data, and its accuracy depends on the size of the training set. In addition, the local structure of the feature points is represented by the change in intensity values of the pixels along a profile line (i.e. edge location) that goes through the feature points. This is based on the assumption that usually the facial features are located on the strong edges. But, finding the correct locations of the feature points on the edges is not always possible.

Ginneken *et al.* [21] proposed to use a non-linear gray level appearance instead of the first derivative profile to model the local structure of the features. In [22], Wei Wang *et al.* had some improvements on the ASM for face alignment. Other authors [23]–[25] used the wavelet transform to model the local structure of features and improve the face alignment. Unfortunately, their approaches are computationally expensive.

The most important limitations of the ASM for facial feature extraction can be summarized as follows.

- 1) Representation of complex multi-part objects by a single shape model. Although a single shape model may preserve the general shape of the whole object, its constraint may fail in extracting some of its parts.
- 2) Representing the local structure of each point by independent models. This drawback may lead to a final shape far from the actual shape model.
- 3) The need for a large training set to cover shape variations. The shape model may fail in characterizing the shape variation if instances of the shape are not incorporated in the training set.
- 4) The initialization of the shape model. This is a major drawback of ASM. If the shape is initialized far from the object of interest, the searching process may either fail or become very slow.
- 5) The choice of modeling the local structure of the points. Many variations of ASM model the local structure by edges or statistical models of gray level variations.
- 6) Search for the best candidate feature points. Most existing algorithms rely on Euclidean or Mahalanobis distance between the candidate feature points and the trained model of the local structure of the feature points.

Some of the above are inherent limitations in ASM and some of them depend on the method used for modeling the local structure of the feature points. For example, the representation of complex multi-part objects by a single shape model is an inherent problem of ASM; while alignment of the shape model to a new instant of the object is not. In this paper, we elaborate on our previous work [26]. Specifically, we deal with model initialization, modeling the local structure of feature points, and shape model alignment to a new instant of the object.

We use the color information to initially detect few salient facial features such as the centers of the mouth, and the eyes. These points are employed to initialize the ASM. We also use the color information to improve the

model that characterizes the local structure of the feature points. A weighted sum of three multivariate Gaussian models for the three components (i.e. Red, Green, and Blue) is used to represent the normalized first derivative of pixel values along a profile line. Furthermore, for the lips, we use the color information to detect their boundary. This enhances the localization of geometric features that represent the external boundary of the lips in face images. In addition, we use 2D affine transformation to align the extracted facial features to the shape model. The 2D affine transformation compensates the effect of head pose variations and the projection of 3D data to 2D. In fact, ASM needs a large training set to cover the variations caused by head pose. For frontal images, the Euclidean transformation is a suitable transformation to align the extracted facial features to the shape model, but for large variations in the head pose, it is not suitable.

The use of the color information for modeling the local structure of the feature points and the use of the 2D affine transformation are general improvements to the ASM approach. On the other hand, the initialization of the ASM using the centers of the mouth and the eyes, and localizing the feature points around the lips are specific for facial feature extraction. Our experimental results show that the proposed approach outperforms the standard ASM technique for facial feature extraction.

The rest of this paper is organized as follows. Section II discusses the improvements of the ASM-based approach for facial feature extraction. Section III demonstrates the proposed algorithm with experimental results and the conclusions are given in Section IV.

II. ENHANCEMENT OF ACTIVE SHAPE MODEL USING COLOR INFORMATION

The ASM approach represents the structure of a target by a parameterized statistical model. By choosing the model parameters, different variations of a target shape can be obtained. In this Section, first we review the theoretical background of the ASM and subsequently we will present our method for improving the technique.

In the ASM technique, the location of n landmark points (e.g. facial features in our work), are annotated on a set of training images by a human expert. This set of points is represented by a vector $X = (x_1, y_1, \dots, x_n, y_n)^T$, where x_i and y_i are the coordinates of the i^{th} landmark. Then, a model that incorporates the variations in shape over the training set is represented as follows:

$$X \approx \bar{X} + P b \quad (1)$$

The vector \bar{X} contains the mean values of the coordinates of the annotated data, P is a matrix of the first t eigenvectors of the covariance matrix of the data, and b is a vector that defines the model parameters. The variance of the i^{th} parameter, P_i , across the training set is given by the corresponding eigenvalue λ_i . By limiting the parameter b_i in the range of $\pm 3\sqrt{\lambda_i}$, we ensure that the generated shape is similar to those in the original training set. To apply the

created shape model to a given target shape, we need to find a transformation to move from the model coordinate system to the image coordinate system. Typically, this is achieved by an Euclidean transformation defining the translation (X_t, Y_t) , rotation θ , and scale s . Therefore, the position of the model points, X , in the image are given by:

$$X = T_{X_t, Y_t, \theta, s}(\bar{X} + P b) \quad (2)$$

For a given new image, the ASM is performed to locate the target object lies in the image. Therefore, we need to find the optimum parameters of the ASM that best fit the model to the structure of the target. Generally, this optimization problem is solved iteratively [1]. At the first step, the model is initialized by the mean shape. Afterward, a region of the image around each feature point is examined to find the best nearby match (i.e. searching along the profile line for the edge locations). In the next step, the parameters X_t, Y_t, s, θ , and b are updated to best fit the new found points. Then, the constraint $|b_i| < 3\sqrt{\lambda_i}$ is applied to the parameters b_i . These steps are repeated until there is no significant change in the shape parameters.

A. Shape model initialization and face alignment using 2D affine transformation

For a 3D object like the face, since it may have some pose variations in the captured images, a simple Euclidean transformation with 4 degrees of freedom is not effective especially when there are large pose variations. When the head is rotated to the left or right, this problem is more severe. To solve this problem, we use a 2D affine transformation with 6 degrees of freedom, given by Equation 3, to align the extracted feature points in the image coordinate system with the points represented by the model coordinate system.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3)$$

To find the parameters of the 2D affine transformation, we need at least three corresponding points, which are not collinear.

As we mentioned in Section I, the initialization of the ASM is very important. With poor initialization, the search process may either fail or become slow. Therefore, a good initialization would help in finding the optimum solution in less iterations. We use our algorithm in [14] to find the centers of the mouth and the two eyes. Figure 1 shows the extracted locations of the eyes and the mouth in a given color face image using this method. We use these three points to obtain the affine parameters in Equation 3 to initialize the shape model for the image.

B. Improvement of the local structure model

In the original ASM, the local structure of a feature point is modeled by assuming that the normalized first derivative of the pixel intensity values along a profile line



Figure 1. Detecting the center of the mouth and the eyes as the salient features for initialization.

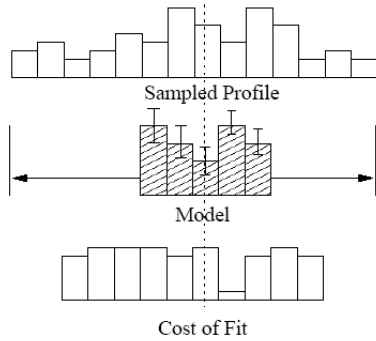


Figure 2. Searching along a sampled profile to find the best fit.

satisfies a multivariate Gaussian distribution. This gives a statistical model for the profile around the point. As shown in Figure 2, a sampled profile, g_s , is matched to a reference model by searching along the profile line that goes through the point and finding the best fit. This is achieved by minimizing the Mahalanobis distance:

$$f(g_s) = (g_s - \bar{g})^T \Sigma^{-1} (g_s - \bar{g}) \quad (4)$$

where \bar{g} is the mean value and Σ is the covariance matrix.

In this paper, we assume that the three color channels, i.e. Red, Green, and Blue, are statistically independent and the normalized first derivative of the color values along a profile line for each individual channel satisfies a multivariate Gaussian distribution. Then, we use a weighted sum of Mahalanobis distances for the three color channels to find the best match for the feature points. Similar to Equation 4, the best matching of a probe sample, g_{rgb} , in RGB color space to a reference model is carried out by minimizing:

$$f(g_{rgb}) = \sum_{i \in r, g, b} w_i (g_i - \bar{g}_i)^T \Sigma_i^{-1} (g_i - \bar{g}_i) \quad (5)$$

where w_i is the weighting factor for the i^{th} component of the Gaussian model with unit sum.

C. Enhancement of the location of features around the lips

To accurately localize the feature points around the lips, we segment the lips region from the facial skin region. This is achieved by classifying each pixel, $I(r,g,b)$,

in RGB color space, either as facial skin or lips. This classification starts by applying the Fisher Discriminant Analysis (FDA) technique. FDA is a data analysis technique that provides more class separability than Principal Component Analysis (PCA), which provides better data representation. By applying FDA to each pixel of the color image, $I(r,g,b)$, we obtain a scalar function that can be used to discriminate between the two classes: facial skin and lips. This function is calculated by using the within-class scatter matrix and is defined as:

$$Fisher(I) = W.I^T \quad (6)$$

where I is a given pixel value. The projection vector W , is calculated by:

$$W = S_W^{-1}(m_1 - m_2) \quad (7)$$

The within class scatter matrix S_W , is:

$$S_W = S_1 + S_2 \quad (8)$$

and

$$S_i = \sum (I - m_i)(I - m_i)^T \quad (9)$$

The sample mean vector of each class, m_i , is defined as:

$$m_i = \frac{1}{n_i} \sum_{I \in D_i} I \quad (10)$$

Where D_i is the set of the pixels in i^{th} class and n_i is the number of the pixels in the class.

To learn the projection matrix W , we obtain a color database of facial images and manually extract patches of lips and facial skin regions. Then, the matrices S_i and m_i are calculated for the two classes of facial skin and lips. For a given test image, we apply the FDA function (Eq. 6) and apply threshold to the result to segment the lips from the facial skin. Figure 3 shows the results of the different steps for lips detection applied to a sample image in our database using FDA. Figure 3(a) is the original image, Figure 3(b) shows the result of applying FDA, Figure 3(c) shows the result of thresholding. The value of the threshold is estimated by a small training set different from the images used in the experiments. The result of applying morphological operators to remove noise and fill holes is shown in 3(d). The following summaries our iterative approach for facial features extraction using enhanced active shape model:

- 1) Extract the centers of the eyes and the mouth using [14].
- 2) Initialize the shape model based on the extracted three points in step 1.
- 3) Calculate the shape parameters, b .
- 4) Examine a region of the image around each point, (x_i, y_i) , to find the best nearby match for that point using the enhanced color model.
- 5) Use lips detection to tune the feature points around the mouth.
- 6) Update the parameters of the affine transformation $(a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)$ to best fit the new found locations of the instance, X , of the target model.

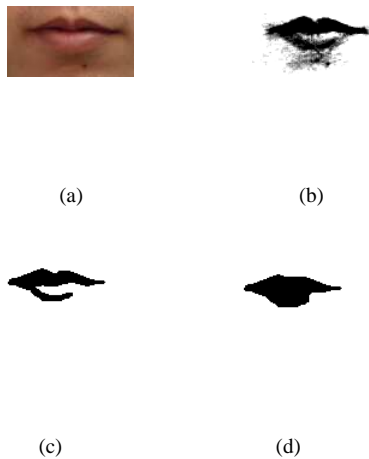


Figure 3. Lips detection using FDA. (a) Original image. (b) Result of FDA classification. (c) Result of thresholding. (d) Result of applying morphological operators.

- 7) Apply the constraints to the parameters, b , to ensure reasonable shapes (i.e. $|b_i| \leq 3\sqrt{\lambda_i}$).
- 8) Go to step 3 and repeat until convergence.

III. EXPERIMENTS AND RESULTS

In this section we validate our algorithm in two ways. First, we compare the error rates of the standard ASM algorithm and our improved ASM approach with respect to the manual annotation. Second, we compare between the two methods by applying face recognition using the extracted facial features. In our comparison, we use a database of 70 different subjects with a total of 140 near frontal color images. For each subject, we captured two near frontal images. For the error rate evaluation, we use one image for each subject. For the face recognition evaluation, one image is used as a probe and another image is used for the gallery. We use a trained shape model for 75 facial feature points which is provided from a completely different source of images and other public databases. Figure 4 shows few samples of the facial images in our database and the extracted facial features by both our method and the original ASM method. The visual inspection shows that our approach is more successful in extracting the locations of the facial feature points, especially around the lips and the corner of the eyes and the eyebrows. From these few examples, it is clear that our method leads to more accurate localization of the facial features as we further prove next.

A. Performance evaluation

To evaluate the performance of our method, we manually labeled 75 feature points in each of the images in the database. For each subject image, the 75 facial feature points were extracted using the original ASM and our improved approach. The performance was evaluated using the average mean square error (MSE) over all the images. The error is defined as the distance between the manually

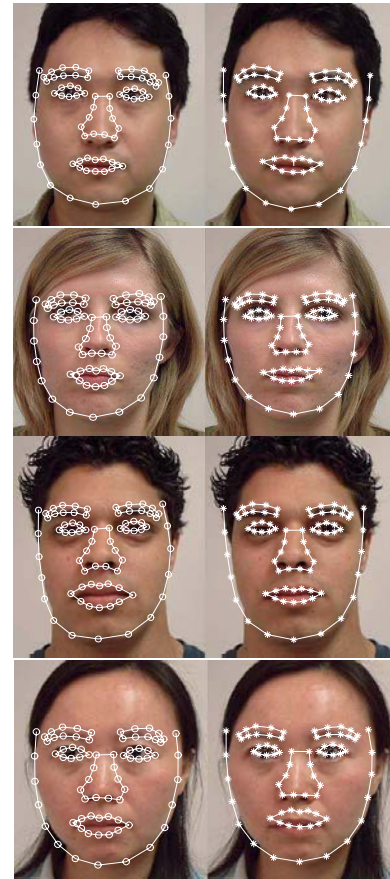


Figure 4. Sample images from our database with extracted features, where our method outperforms the standard ASM method, (o) Improved ASM, (*) Standard ASM. The MSE for the extracted features using both the enhanced ASM and the standard ASM with respect to the manually labeled features for these sample images are (a) [28.85, 39.98], (b) [32.74, 33.54], (c) [20.50, 38.85], and (d) [39.29, 110.87].

labeled feature points and the corresponding feature points obtained from both the original and our improved version of the ASM approach. This is defined as follows:

$$AverageMSE = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{n} \sum_{j=1}^n \|P_{ij} - P'_{ij}\|^2 \right) \quad (11)$$

where N is the total number of images, n is the number of the landmark points in the shape model, P_{ij} is the j^{th} landmark point in the manually labeled shape of the i^{th} test image, P'_{ij} is the j^{th} landmark point in the resulting shape of ASM for the i^{th} test image, and $\|\cdot\|$ denotes the Euclidian distance. Figure 5 shows a plot of the MSE of the two methods over all the 70 subjects in the database, on a case by case based. We can see that our approach generally has better performance for most of the cases. As given in table II, we categorize our results in two sets based on the average MSE. In one set, 49 subjects out of the 70 subjects, the average MSE between the manually labeled feature points and the feature points extracted by our improved ASM are lower than the average MSE between the manually labeled feature points and the feature points extracted by the

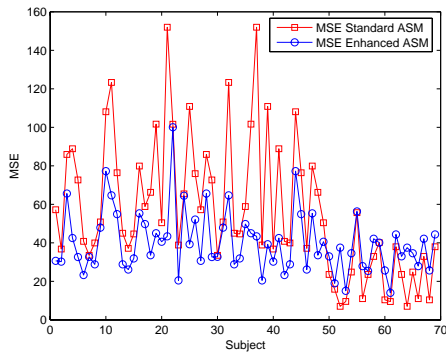


Figure 5. The Mean Square Error for the standard ASM and the enhanced ASM.

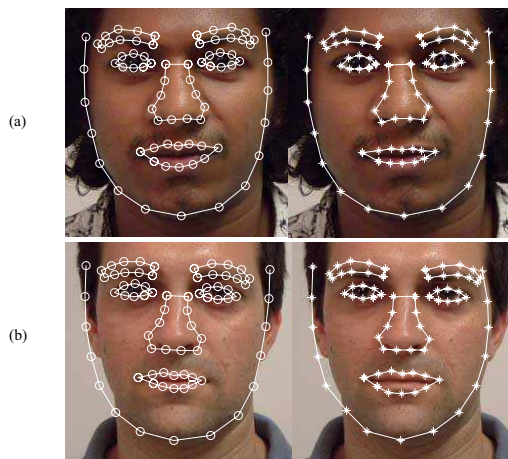


Figure 6. Sample images from our database with extracted features, where the standard ASM outperforms our method, (o) Improved ASM, (*) Standard ASM. The MSE for the extracted features using both the enhanced ASM and the standard ASM with respect to the manually labeled features for these sample images are (a) [34.64, 24.87] and (b) [14.02, 9.55].

original ASM method. For the second set, the remaining 21 subjects, the original ASM has lower average MSE. In other words, for 70% of the images in the database, our improved ASM has lower average MSE than the standard ASM and for the remaining 30% of the images in the database, the standard ASM has lower average MSE than our method.

Figure 4 shows sample images, out of the 70% cases, along with the MSE errors in which our method outperforms the standard ASM approach with minimum error. Similarly, Figure 6 gives sample cases where the standard ASM performs better than our method. In these sample examples, the figure visually shows comparable performance between the two methods yet based on the MSE, the standard ASM attains lower error in 30% of the cases.

TABLE II.
PERFORMANCE COMPARISON BETWEEN THE TWO METHODS AND THE MANUALLY LABELED FEATURES BASED ON THE AVERAGE MSE OVER ALL 70 SUBJECTS.

Method	Ave.MSE for 49 subjects out of 70	Ave. MSE for 21 subjects out of 70	% of cases which have minimum error
Standard ASM	70.3	23.3	30.0%
Enhanced ASM	40.0	31.9	70.0%

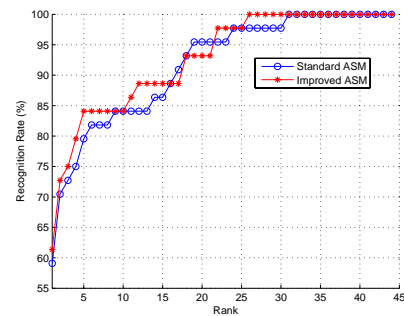


Figure 7. Face recognition using the extracted facial features by the improved ASM and the standard ASM.

B. Face recognition

In this experiment, we compare the results of face recognition using the automatically extracted feature points from our approach and from the standard ASM approach. A set of images was used for probes and another set was used for the database. To recognize a face based on the facial feature points, we align (scale, rotate, and transform) the facial feature points of a given probe image to the feature points of each image in the face database and then calculate the Euclidian distance between them. This process is done for all of the images in the face database. Then, we rank the candidates based on the calculated Euclidian distances. Figure 7 compares the results of face recognition based on the location of feature points extracted by the two techniques. The Figure shows that the features obtained with the enhanced ASM version lead to better recognition than those obtained from the standard ASM.

IV. CONCLUSIONS

In this paper we presented an improved version of the Active Shape Model approach for facial feature extraction. We use the color information to localize the centers of the mouth and the eyes to improve the initialization step in the standard ASM. In addition, we model the local structure of the feature points in the RGB color space and we use a 2D affine transformation to align the facial features that are perturbed by head pose. Experiments show that our improved version of the ASM is accurate and outperforms the standard ASM.

REFERENCES

- [1] T. Cootes, D. Cooper, C. Taylor, and J. Graham, "Active shape models - their training and application." *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
- [2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, December 2003.
- [3] S. Kobayashi and S. Hashimoto, "Automated feature extraction of face image and its applications," in *International Workshop on Robot and Human Communication*, 1995, pp. 164–169.
- [4] F. G. D. Natale, D. D. Giusto, and F. Maccioni, "A symmetry-based approach to facial features extraction," in *13th International Conference on DSP*, vol. 2, July 1997, pp. 521–525.
- [5] A. Nikolaidis, C. Kotropoulos, and I. Pitas, "Facial feature extraction using adaptive hough transform, template matching and active contour models," in *13th International Conference on DSP*, vol. 2, July 1997, pp. 865–868.
- [6] C. Lau, W. Cham, H. Tsui, and K. Ngan, "An energy function for facial feature extraction," in *International Symposium on Intelligent Multimedia, Video and Speech Processing*, vol. 1, 2001, pp. 348–351.
- [7] G. Yen and N. Nithianandan, "Facial feature extraction using genetic algorithm," in *Congress on Evolutionary Computation*, 2002.
- [8] K.-H. Seo, W. Kim, C. Oh, and J.-J. Lee, "Face detection and facial feature extraction using color snake," in *IEEE International Symposium on Industrial Electronics*, vol. 2, 2002, pp. 457–462.
- [9] D. Xi, , and S. Lee, "Face detection and facial component extraction by wavelet decomposition and support vector machines," in *4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, 2003, pp. 199–207.
- [10] K.-A. Kim, S.-Y. Oh, and H.-C. Choi, "Facial feature extraction using pca and wavelet multi-resolution," in *Face and Gesture Recognition*, 2004, pp. 439–444.
- [11] K.-M. Lam and Y.-L. Li, "An efficient approach for facial feature detection," in *4th International Conference on Image Processing*, vol. 2, 1998.
- [12] Y. Yagi, "Facial feature extraction from frontal face image," in *International Conference on Signal Processing*, 2000.
- [13] Z. Xue, S. Z. Li, D. Shen, and E. K. Tsoh, "A novel bayesian shape model for facial feature extraction," in *Seventh International Conference on Control, Automation, Robotics And Vision (ICMCV02)*, 2002.
- [14] R. Hsu, M. Abdel-Mottaleb, and A. Jain, "Neural network-based face detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, May 2002.
- [15] Y. Hu, B. Yin, and D. Kong, "A new facial feature extraction method based on linear combination model," in *International Conference on Web Intelligence*, 2003.
- [16] A. Gunduz and H. Krim, "Facial feature extraction using topological methods," in *International Conference on Image Processing (ICIP)*, 2003, pp. I: 673–676.
- [17] K. Nagao, "Bayesian approach with nonlinear kernels to feature extraction," in *17th International Conference on Pattern Recognition (ICPR04)*, 2004.
- [18] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 2, no. 1, pp. 91–108, 1996.
- [19] A. K. Jain, Y. Zhang, and M. Dubuisson-Jolly, "Deformable template models: A review," *Journal of Signal Processing*, vol. 71, no. 2, pp. 109–129, 1998.
- [20] A. Lanitis, C. Taylor, and T. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, July 1997.
- [21] B. V. Ginneken, A. Frangi, J. Staal, B. T. H. Romeny, and M. Viergever, "A non-linear gray-level appearance model improves active shape model segmentation," in *IEEE workshop on Mathematical Models in Biomedical Image Analysis*, 2001, pp. 205–212.
- [22] W. Wang, S. Shan, W. Gao, B. Cao, and B. Yin, "An improved active shape model face alignment," in *International Conference on Multimedia*, 2002, pp. 523–528.
- [23] B. Zhang, W. Gao, S. Shan, and W. Wang, "Constraint shape model using edge constraint and gabor wavelet based search," in *4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, 2003, pp. 52–61.
- [24] F. Jiao, S. Li, H. Shum, and D. Schuurmans, "Face alignment using statistical models and wavelet features," in *Computer Vision and Pattern Recognition*, 2003, pp. I: 321–327.
- [25] C. Hu, R. Feris, and M. Turk, "Active wavelet networks for face alignment," in *British Machine Vision Conference*, 2003.
- [26] M. Mahoor and M. Abdel-Mottaleb, "Facial features extraction in color images using enhanced active shape model," in *Seventh International Conference on Automatic Face and Gesture Recognition*, 2006.

Mohammad H. Mahoor received the B.Sc. degree in electrical engineering from Petroleum University of technology, Iran, in 1996 and the M.Sc. degree in biomedical engineering from Sharif University of technology, Iran, in 1998. He is currently working toward his Ph.D. degree in department of Electrical & Computer Engineering, University of Miami. His research interests are pattern recognition, image processing and computer vision.

Mohamed Abdel-Mottaleb received his Ph.D. in computer science from University of Maryland, College Park, in 1993. He is an associate professor in the department of Electrical and Computer Engineering, University of Miami, where his research focuses on 3D face recognition, dental biometrics, visual tracking, and Human activity recognition. Prior to joining the University of Miami, from 1993 to 2000, he worked at Philips Research, Briarcliff Manor, NY. At Philips Research, he was a Principal Member of Research Staff and a Project Leader, where he led several projects in image processing, and content-based multimedia retrieval. He holds 20 US patents and published over 70 papers in the areas of image processing, computer vision, and content based image retrieval. He is an associate editor for the Pattern Recognition journal.

A-Nasser Ansari received the Bachelor of Science degree in Electrical Engineering from the University of Miami, Coral Gables, Florida, in 1990, the Master of Science degree in Electrical and Computer Engineering from the University of Miami, Coral Gables, Florida, in 1994. He is currently pursuing his Ph.D. in Electrical and Computer Engineering. His research interests are image processing, machine vision, and pattern recognition.