An Effective Initialization for ASM-Based Methods

Hong-Quan Hua[1], T. Hoang Ngan Le[2], and Bac Le[1]

[1] Faculty of Information Technology, VNUHCM, University of Science, Ho Chi Minh, Vietnam
[2] Department of Electrical and Computer Engineering (ECE), Carnegie Mellon University, Pittsburgh, USA

Abstract. Locating facial feature points is an important step for many facial image analysis tasks. Over the past few years, Active Shape Model (ASM) has become one of the most popular approaches to solve this problem. However, ASM-based methods are sensitive to initialization errors caused by poor face detection results. In this paper, an effective initialization for the ASM-based methods is proposed by our improved initial 8-landmarks ASM model together with Viola-Jones eye detection. In particular, we apply the 8-landmarks ASM model with the position of eyes from eye detector as reference points. After that, we choose several 76-landmark candidates from the training set that have the key feature points related to the result of the previous 8-landmark model. The best candidate has the lowest fitting errors with the test image and is used as initialization. To evaluate the performance of our work, we conduct the experiments on the MUCT and LFPW database. Compared to the latest ASM implementation, MASM, our proposed can improved the MASM by an average accuracy of 44% when dealing with poor initialization.

Keywords: Active Shape Models, Facial feature points, Initialization errors.

1 Introduction

Facial feature points play an important role for face analysis in many computer vision systems such as human-machine interface, face recognition and expression analysis. For this reason, the problem of facial feature points localization has attracted large amount of attention. As shown in the recent research by Chen, et al.'s [2], leading performance in face verification can be acquired with simple features if facial feature points are detected accurately; therefore the problem of feature localization has become significant and attracted in the past years. Many approaches to extract facial feature points has been studied extensively including the active contour [10], deformable template [19], Active Shape Model [5], Active Appearance Models [4]. Most of above methods combine the local texture around each point and the shape of the whole face. ASM is known as one of the most popular methods that finds the landmarking points by Principle Component Analysis (PCA).

K. Saeed and V. Snášel (Eds.): CISIM 2014, LNCS 8838, pp. 421–432, 2014.

Active Shape Models (ASM) [5] proposed by Cootes, et al. has achieved great success in extracting facial feature points. In ASM, the feature points are modeled by the gray-scale texture whereas the geometrical information is defined by PCA. In order to implement a ASM system, a face detection is first used to find face regions and initialize a starting shape, then ASM is used to refine feature positions to the best locations based on the learned texture model. ASM is like deformable methods like snakes [10], but ASM's feature points are constrained to be a valid shape by the learned statistical shape model. Though many methods have been proposed to improve the fitting accuracy of ASMs [7],[13], [9], facial feature points localization remains a very challenging problem. The challenge comes from the variations of facial appearance in illumination, pose, expression, aging and so on. Especially, ASM methods are very sensitive to the initialization perturbations (translation, rotation and scaling in size) caused by the poor face detection [16]. To overcome these defects, Cootes et al. [3] propoesd a multi-resolution method, Wang et al. [18] added edge information into the model. Milborrow and Nicolls [12] achieved fairly accurate fitting results by combining 1D and 2D search model in the multi-resolution approach. Seshadri and Savvides [15] proposed a new robust method, called Modified Active Shape Model (MASM). MASM contains many theoretical changes to the traditional ASM and gains better results than above methods. However, MASM is still sensitive to initialization errors, especially while dealing with in-plane rotations and scaling effects. This is because the faces used to train the methods were upright and the initialization is too far from the optimal positions.

In our work, we focus on a new method to overcome the poor initialization. We select the initialization based on trained shapes instead of using mean shape as starting shape. Our selection process takes advantage of eye detector, support vector machine. Our selected feature points are almost in the optimal positions, which can prevent problems due to local minima. We also inherit the optimization metric of MASM [15] to determine the best location for each feature point. However, we weaken the important of edge information while fitting the facial boundary.

The remainder of this paper is organized as follows. In Section 2, the basic of Active Shape Model and improvements in Modified Active Shape Model are introduced. Our proposed method is described in Section 3. Experimental results are given in Section 4. A conclusion and future work are drawn in Section 5.

2 Related Works

This section gives an overview of Active Shape Models and the latest ASM-based methods called MASM. First, we explain the basic of training and testing stage in ASM method. Then, we show improvements of MASM to the basic ASM model since we use it as our main ASM-based method.

2.1 Active Shape Model

The following gives a brief review of the original ASM containing two main procedures: Training Stage and Testing Stage.

Training Stage. Shape Model and Local Appearance Model are built from a training set of manually annotated images. The Local Appearance Model is used to optimize the location for each landmark. On the other hand, the Shape Model constrains the landmark points to be in a valid shape.

The Shape Model: A *landmark* represents a distinguishable feature point in facial images. Locating landmarks is the same as locating feature points. Each face shape can be described by the coordinates of N landmarks points. The landmark points are (manually) annotated by a set of K training images. In our experiment, we use MUCT face database [11] which contains a total of 3755 faces with 76 manual landmarks at the resolution of 640 x 480 pixels. Figure 1 shows some example images of MUCT database.

Fig. 1. Example images from MUCT database. The green dots indicate the manually labeled landmark points.

The coordinates of each landmarks for each image are stacked as a shape vector $\mathbf{x}$.

$$\mathbf{x} = [x_1 \ x_2 ... x_N \ \ y_1 \ y_2 ... y_N]^T \tag{1}$$

where x_i and y_i are the x and y coordinates of the ith landmark and N is the number of landmarks used.

The training images shape are all aligned by using Generalized Procrustes Analysis (GPA) [8] for minimizing the sum of squared distances between corresponding landmark points. PCA is then applied to reduce the dimensionality of the shape vectors, such that any shape $\mathbf{x}$ can be approximated as follows.

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{Pb} \tag{2}$$

Here $\bar{\mathbf{x}}$ is a mean shape vector, $\mathbf{P}$ is a matrix of eigenvectors and $\mathbf{b} = (b_1 \ b_2 ... b_t)^T$ is a vector of shape parameters. When fitting the shape model to a face, the value of b_i is constrained to lie with the range ± 3 standard deviations to ensure that generated face shapes are valid.

The Local Appearance Model: The statistical models of appearance models (grayscale pixel intensities) of local regions around each landmark is used to optimize the location for each landmark in the search process. The 1D profiles are constructed by sampling the pixel intensities along a line, which is perpendicular to the landmark contour and centered at that landmark. The length of this line is $2k + 1$ with k pixels sampled on either side of the landmark. The normalized derivative of this vector is computed to form the profile vector, $\mathbf{g}$. The mean of such profiles (over the training images) at each landmark is called the mean profile vector, $\bar{\mathbf{g}}$ and the covariance matrix of such vectors is $\mathbf{S_g}$. To improve the accuracy and robustness, the process is repeated in a multi-resolution framework. The pyramid image contains L_{max} levels, where image at level i is formed by smoothing the image at level $i-1$ and scaling by 50 percent. The fitting process is carried out at the coarse level first, then the results is refined in the finer level.

To compare the difference between a new sampled profile and the mean profile, we use the Mahalanobis distance measure as follows:

$$f(\mathbf{g_i}) = (\mathbf{g_i} - \bar{\mathbf{g}})^T \mathbf{S_g^{-1}} (\mathbf{g_i} - \bar{\mathbf{g}}) \tag{3}$$

By minimizing $f(\mathbf{g_i})$, we can get the optimal position for the landmark.

At each landmark position, a 2D profile is constructed by sampling a local patch around the landmark. The center of this local patch is the current landmark. The 2D local patch is then vectorized and normalized into vector $\mathbf{g}'$ using a constant parameter c at each element of the profile, g_i as shown in (4).

$$g_i' = \frac{g_i}{g_i + c} \tag{4}$$

where g_i' is the ith element of the vector $\mathbf{g}'$

Testing Stage. The search process of ASM can be implemented two iterative steps: (1) update the best location for each landmark point, (2) update the shape parameter b to ensure a valid shape. In the first step, we need to sample m pixels on both side of each landmark ($m > k$). We then compute the Mahalanobis distance as above for each sub-profile with the length of k. The sub-profile with the minimum distance is chosen and the the center of that sub-profile is the update position for the current landmark. Once all landmarks have been moved to the best position, we obtain the new shape vector $\mathbf{x_{ini}}$ by using the shape model. The parameter vector $\mathbf{b}$ is calculated by minimizing the cost function in (5)

$$|\mathbf{x_{ini}} - T(\bar{\mathbf{x}} + \mathbf{Pb})|^2 \tag{5}$$

In (5), b is the vector of shape parameter, T is the similarity transform that minimizes the Euclidean distance between $\mathbf{x_{ini}}$ and $\bar{\mathbf{x}} + \mathbf{Pb}$. The detail of the process to find b and T can be found in [6].

The adjusting process is repeated until no significant change in the landmark points in observed. Once convergence is reached at the lowest resolution level, the landmark coordinates is scaled and the entire process is repeated in lower level until convergence at the finest level. Figure 2 gives overview of ASM steps at testing stage.

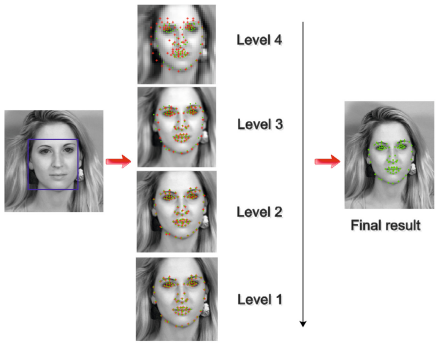


Fig. 2. Overview of ASM steps at testing stage. The blue box is the result of face detector. Red dots indicate the starting shape for each step and Green dots indicate the landmark result.

2.2 Modified Active Shape Model

In this section, we describe the Modified Active Shape Model (MASM) proposed by Keshav Seshadri and Marios Savvides [15]. This approach gives significant improvements to the traditional ASM. MASM can handles faces with slight pose variations, in-plane rotations and varied expressions.

Firstly, MASM uses an optimal number of landmark points to represent face shapes. The landmarking scheme used in MASM contains 79 points which can perform accurately and model varied facial expressions. Secondly, edge detection is used to better fit points around the facial boundary. The main contribution of MASM is the development of new metric to determine the best location of the landmarks compared to the minimum Mahalanobis distance in traditional method.

MASM metric can be described as follows. MASM extracts 2D profiles and builds a subspace to model the pixel variation in the local patch around each landmark. They obtains the mean profile vector for each landmark and computes the eigenvectors of the covariance matrix $\mathbf{S_g}$ corresponding to the dominant eigenvalues that models 97% of variation. Eigenvectors are stored along the columns of matrix $\mathbf{P_g}$. In the testing stage, a profile $\mathbf{g}$ around a candidate landmark is projected onto the subspace using equation (6) to obtain the shape parameter vector $\mathbf{b_g}$ and is then reconstructed using this parameter vector to obtain $\mathbf{g_r}$ using equation (7).

$$\mathbf{b_g} = \mathbf{P_g}^T(\mathbf{g} - \bar{\mathbf{g}}) \tag{6}$$

$$\mathbf{g_r} = \bar{\mathbf{g}} + \mathbf{P_g}\mathbf{b_g} \tag{7}$$

The metric for determining the best candidate point is the Mahalanobis distance between the original profile $\mathbf{g}$ and the reconstructed profile $\mathbf{g_r}$. The candidate with lowest reconstruction error is the best fit because it is the one that

can be best modeled by the subspace. The reconstruction error can be computed as in equation (8).

$$error(\mathbf{g}, \mathbf{g_r}) = (\mathbf{g_r} - \mathbf{g})^{\mathbf{T}} \mathbf{S_g^{-1}} (\mathbf{g_r} - \mathbf{g}) \tag{8}$$

MASM also use edge information to improve the fitting accuracy for facial boundary. They assume that landmark points along the facial boundary usually lie along strong edges. The error function for these landmark is now defined in equation (9).

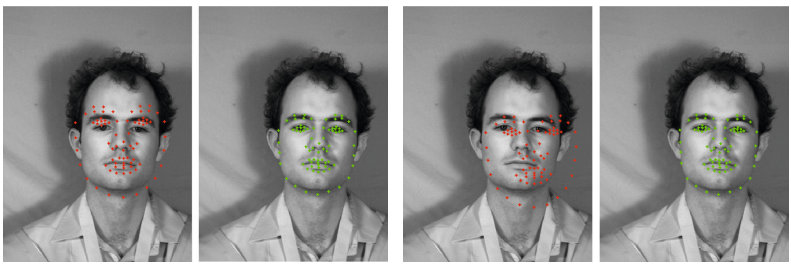
$$error(\mathbf{g}, \mathbf{g_r}) = (c - I)(\mathbf{g_r} - \mathbf{g})^{\mathbf{T}} \mathbf{S_g^{-1}} (\mathbf{g_r} - \mathbf{g}) \tag{9}$$

where c is a constant, chosen to be 2 in MASM implementation and I is the Sobel edge intensity at the candidate point which can only take on the values 1 or 0.

These modifications allow MASM to gain a significant improvement over the traditional ASM model.

3 Our Proposed Method

In this section, we introduce our proposed method for choosing initialization. In our method, each landmark is lie in an optimal location even when slightly expression and initialization errors are present. Moreover, we also introduce a modification to the cost function of MASM to improve the fitness of facial boundary. Figure 3 shows the comparison of initialization between our proposed method and MASM.



(a) (b)

Fig. 3. Initialization of MASM and our proposed method without initialization errors a) and with initialization errors b). The Red dots indicate the initialization of MASM. The Green dots indicate the initialization of our proposed method.

3.1 New Method for Choosing Initialization

Most of previous ASM methods can not recover from poor initialization. The initialization is usually provided by the bounding box around the face which is the result of face detector. However, face detection algorithms such as Viola-Jones face detector [17] have no guarantees that the localization is perfect. There are many errors in translation and size. Especially, the bounding box from face detector is usually upright so the localization is poor when dealing with in-plane rotation of a face.

To overcome this problem, we propose a new method to reduce the sensitivity of ASM model to poor initialization. In our method, we use the location of eyes, nose and mouth to find 20 shape candidates from the training test that have the geometrical structure similar with the testing image. After that, we will use the local appearance information around each landmark to decide the best shape candidate. This shape candidate will be used as initialization instead of mean shape in previous ASM methods. Since we use the eyes, nose and mouth locations as well as the local appearance information, the resulting initial landmarks are almost in their optimal locations. The performance of this approach is shown in Section 4.

There are two different ASM model in our method, *full-shape model* and *minimal-shape model*. *Full-shape model* contains all the landmark points as in training set. On the other hand, *minimal-shape model* only contains 8 landmark points including 4 for eyes corners, 2 for nose and 2 for mouth corners. Figure 4 shows examples of both models. A flowchart of the overall process is depicted in Figure 5. First, we use Viola-Jones face detector to extract the face regions from the test image. Then, we detect the eyes using Viola-Jones algorithm. Since the detected regions contain many false positives, we apply a simple linear Support Vector Machine (SVM) to select true positive eye regions. After that, we apply the *minimal-shape model* with eyes positions as reference points. The result is the optimal positions for 8 landmark points and the corresponding shape parameter vector $\mathbf{b_{min}}$. The key idea is that we will use this $\mathbf{b_{min}}$ to select 10 nearest shape parameter vector of known minimal shape in training set. This means that we now have 10 candidates of full shape for initialization purpose.

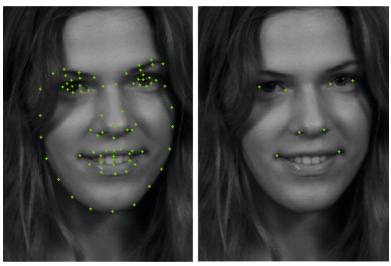


Fig. 4. Examples of *full-shape model* on the left and *mini-shape model* on the right

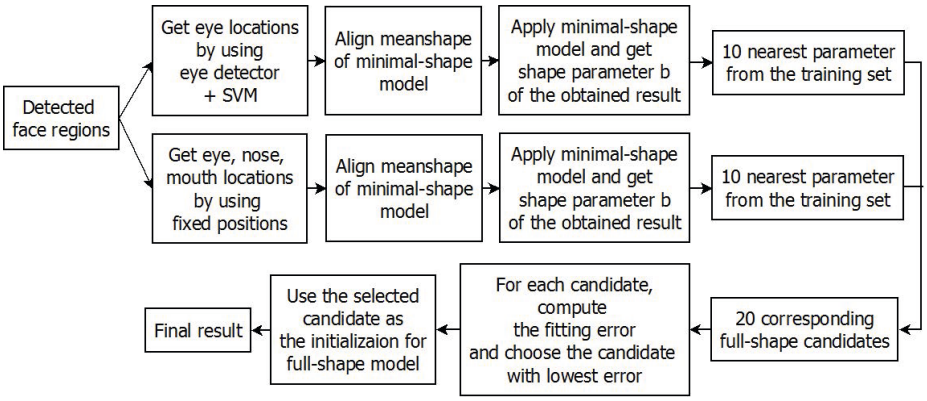


Fig. 5. Sequence of operations to select the suitable initialization for MASM model

Because the eye detector doesn't guarantee perfect results, we repeat the above process to have 10 more candidates of full shape. However, we use fixed positions in the bounding box as reference points for *minimal-shape model* instead of eyes positions. The ratio for fixed positions is described in Figure 6.

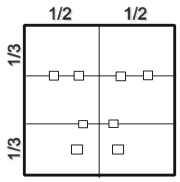


Fig. 6. The position of fixed *minimal-shape* points in the face region

After obtaining 20 candidates for initialization, we will select the candidate that has lowest error value. The error value of each candidate is the summation of fitting error of the profile around each landmark point. The selected candidate will be use as initialization points for *full-shape model*.

3.2 Use of Edge Information

In MASM, Keshav Seshadri and Marios Savvides assume that facial landmark usually lie along strong edges. However, we suggest that we should weaken that assumption because Sobel edge detector don't guarantee that strong edges will lie on facial boundary. In that case, the candidate points may be chosen inaccurately and then the performance of fitting process is reduced.

For the landmarks in the facial boundary, we will use the cost function as in equation (10) to measure Mahalanobis distance between the original profile $\mathbf{g}$ and the reconstructed profile $\mathbf{g_r}$.

$$error(\mathbf{g}, \mathbf{g_r}) = (1 - c * I)(\mathbf{g_r} - \mathbf{g})^{\mathbf{T}} \mathbf{S_g^{-1}} (\mathbf{g_r} - \mathbf{g}) \tag{10}$$

where c is a constant, chosen to be 0.2 in our implementation, I is the Sobel edge intensity at the candidate point and $\mathbf{S_g}$ is the covariance matrix of the current landmark's subspace.

4 Experiments and Results

In this section, we will evaluate the performance of our proposed method and show the experimental results. The comparison experiments between our method and the Modified Active Shape Model (MASM) are conducted on MUCT [11] and LFPW [1] database. All model are trained with 300 face images from MUCT database. The landmark scheme used in MUCT database contains 76 landmark points as in Figure 4. Some original sample images of MUCT database are shown in Figure 1. We conducted two experiments in testing stage. In the first test, we evaluate with two test set from MUCT and LFPW with the assumption that initialization errors were not present. The first set consisted of 450 images from the MUCT database. The second set consisted of 300 images from LFPW database. In the second experiment, we added various of initialization errors to the face detector results.

To evaluate the performance of the method, we use the distance between the predicted landmark positions and the manually annotated points.

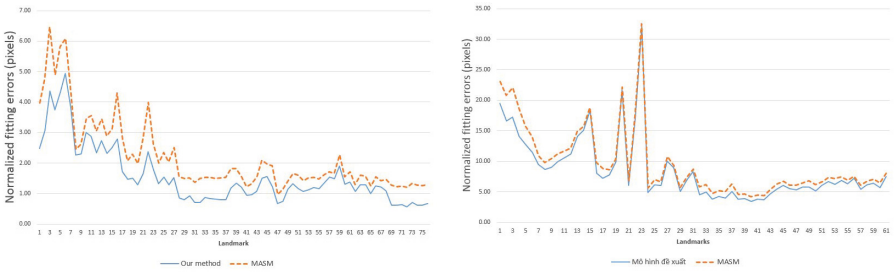
$$E = \frac{1}{n} \sum_{i=1}^{n} \{ \frac{1}{N} \sum_{j=1}^{N} dist(\mathbf{s_{ij}}, \mathbf{m_{ij}}) \} \tag{11}$$

where n is the number of test images, N is the number of landmark, $\mathbf{s_{ij}}$ and $\mathbf{m_{ij}}$ are the coordinate vectors of the automatic landmark and the manually labeled landmark respectively. The Euclidean distance $dist(\mathbf{s_{ij}}, \mathbf{m_{ij}})$ between $\mathbf{s_{ij}}$ and $\mathbf{m_{ij}}$ is normalized so that the distance between the nearest corners of both eyes is 30 pixels. For these experiment, we used Viola-Jones face detector to get the face bounding box. Then, we refined manually the incorrectly detected results in translation and scaling to prevent initialization errors.

The results of first test are illustrated in Table 1. We can find that our method can enhance the accuracy of the MASM by 23.66 % in MUCT test and 14.59% in LFPW test. Figure 7 shows plots the errors for each landmark. The landmark scheme of MUCT database is 76 points. However, the ground truth landmark points in LFPW is 68 points, as labeled by IBUG [14]. Therefore, we choose the corresponding landmark points in both scheme to compute fitting errors. The resulting scheme contains 61 points. As can be seen, our proposed method has better result than the MASM method, especially while fitting the landmarks

Table 1. Results for first test without initialization errors

	Fitting Errors		Improvement
	Our method	MASM	
MUCT	1.5777	2.0666	23.66%
LFPW	7.6756	8.9873	14.59%

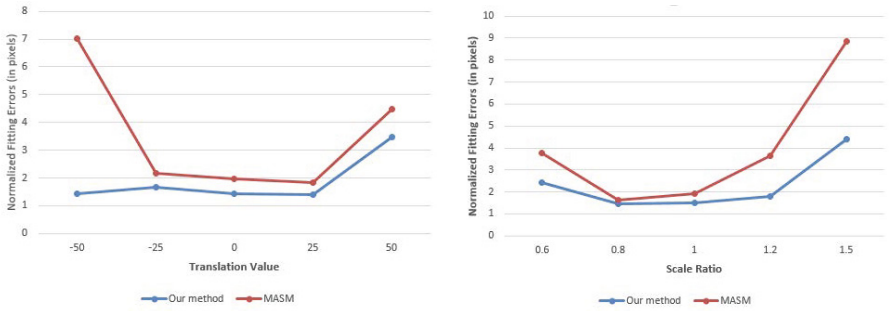


(a) (b)

Fig. 7. Plots of fitting errors vs landmark number for (a) MUCT test (b) LFPW test

along the facial boundary (1-15). This is the effect of the modification in using edge information. On the remaining landmark points, our method just gives slightly improvements because we focus on the problem of initialization errors while these experiments assume that there is no initialization errors. Besides, other processing steps are almost the same as MASM. Hence, we don't have much improvement on others part. In the LFPW test, the errors are higher than the first test because LFPW images are captured in real-world scenario which contain many pose changes and expressions. However, our method can gain better results than MASM.

For our second experiments, we estimate the performance of our method when dealing with initialization errors. In this experiments, we test the effect of translation and scaling to the accuracy of our proposed method and MASM. Figure 8 shows the results of our test. In the translation test, we move the face bounding box with a same value along x and y coordinates. The translation value is: -50, -25, 0, 25, 50. In the scaling test, we scale the face bounding box with ratios of [1.5, 1.2, 1, 0.8, 0.6]. In Scaling test, we achieve the fitting errors of 2.32048 compared to 3.9663 of the MASM. In translation test, the error of our method is 1.8742 while MASM's error is 3.4945. In both cases, our proposed method can gains better performance, 46.37% and 41.49% in translation and scaling test respectively , which means that our method are more robust to the initialization errors.



(a) (b)

Fig. 8. Plots of fitting errors with the effects of scaling a) and translation b)

5 Conclusion and Future Work

In this paper, we presented a novel Active Shape Model for facial feature points detection in frontal-view face images. Our method focus on dealing with initialization errors due to the failure of face detector. In order to recover from poor initialization, we use an eye detector with Support Vector Machine to effectively extraction the location of the eye. Then, we apply a minimal ASM model with 8 key landmark points to obtain an optimal initialization. This optimal initialization is used as the starting point for the full ASM model with 76 landmark points instead of using mean shape. Moreover, we consider the usage of edge information carefully to prevent misalignment on facial boundary. Experimental results show that our proposed method can achieve better fitting accuracy even with poor initialization. In the future work, there are still many rooms to improve the performance of our proposed method. First, our method uses the eye location to obtain the initialization. Hence, our method can't deal with occlusion effectively, especially when the target face has sun glasses. Such problems can be solve by combing the location of nose and mouth in addition to eye location. Secondly, the feature descriptor is very simple. We can improve it with HOG or Gabor features which are robust to illumination.

Acknowledgments. This research is supported by research funding from Advanced Program in Computer Science, University of Science, Vietnam National University - Ho Chi Minh City.

References

1. Belhumeur, P.N., Jacobs, D.W., Kriegman, D., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 545–552 (2011)

2. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3025–3032. IEEE (2013)
3. Cootes, T.F., Taylor, C.J., Lanitis, A.: Multi-resolution search with active shape models. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, onference A: Computer Vision & Image Processing, vol. 1, pp. 610–612. IEEE (1994)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (2001)
5. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
6. Cootes, T.F., Taylor, C.J.: et al.: Statistical models of appearance for computer vision (2004)
7. Cristinacce, D., Cootes, T.F.: Boosted regression active shape models. In: BMVC, pp. 1–10 (2007)
8. Gower, J.C.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
9. Gu, L., Kanade, T.: A generative shape regularization model for robust face alignment. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 413–426. Springer, Heidelberg (2008)
10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
11. Milborrow, S., Morkel, J., Nicolls, F.: The muct landmarked face database. Pattern Recognition Association of South Africa 201(0) (2010)
12. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
13. Romdhani, S., Gong, S., Psarrou, A., et al.: A multi-view nonlinear active shape model using kernel pca. In: BMVC. vol. 10, pp. 483–492 (1999)
14. Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: A semi-automatic methodology for facial landmark annotation. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 896–903. IEEE (2013)
15. Seshadri, K., Savvides, M.: Robust modified active shape model for automatic facial landmark annotation of frontal faces. In: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, BTAS 2009, pp. 1–8. IEEE (2009)
16. Seshadri, K., Savvides, M.: An analysis of the sensitivity of active shape models to initialization when applied to automatic facial landmarking. IEEE Transactions on Information Forensics and Security 7(4), 1255–1269 (2012)
17. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
18. Wang, W., Shan, S., Gao, W., Cao, B., Yin, B.: An improved active shape model for face alignment. In: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, p. 523. IEEE Computer Society (2002)
19. Yuille, A.L.: Deformable templates for face recognition. Journal of Cognitive Neuroscience 3(1), 59–70 (1991)