

# Robust Three-step facial landmark localization under the complicated condition via ASM and POEM

**Weisheng Li<sup>1</sup>, Lai Peng<sup>1</sup>, Lifang Zhou<sup>1</sup>**

<sup>1</sup>Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

[e-mail: liws@cqupt.edu.cn]

[e-mail: ivyicecream@163.com]

[e-mail: zhoulf@cqupt.edu.cn]

\*Corresponding author: Weisheng Li

*Received January 20, 2015; revised May 19, 2014; revised June 20, 2015; accepted July 16, 2015; published September 30, 2015*

---

## Abstract

To avoid influences caused by pose, illumination and facial expression variations, we propose a robust three-step algorithm based on ASM and POEM for facial landmark localization. Firstly, Model Selection Factor is utilized to achieve a pose-free initialized shape. Then, we use the global shape model of ASM to describe the whole face and the texture model POEM to adjust the position of each landmark. Thirdly, a second localization is presented to discriminatively refine the subtle shape variation for some organs and contours. Experiments are conducted in four main face datasets, and the results demonstrate that the proposed method accurately localizes facial landmarks and outperforms other state-of-the-art methods.

---

**Keywords:** Facial Landmark Localization; Active Shape Model (ASM); Model Selection Factor (MSF); Patterns of Oriented Edge Magnitudes (POEM); Pattern Recognition

---

This research was supported by the Natural Science Foundation of China (No.61272195, 61472055, 61100114, U1401252), Program for New Century Excellent Talents in University of China (NCET-11-1085), Chongqing Outstanding Youth Found (cstc2014jcyj40001), Chongqing University of Posts and Telecommunications student research training program (A2014-07) and Chongqing Research Program of Application Foundation and Advanced Technology (cstc2012jjA40036).

## 1. Introduction

Face recognition [1, 2], as the significant pattern recognition object, has recently been widely studied and applied in many fields [3, 4]. In general, the process of face recognition involves three key phases: face detection, feature localization and extraction, classification [5-7] and identification. Facial landmark localization or facial feature point localization, defined as the localization of certain feature points on the face, is a foundational and essential task in the second phase of face recognition. The model-based methods [8-10], considering that the whole face and the ensemble of landmarks as an instantiation of a shape, are one of the major ways to localize facial landmarks. Most of them mainly include topics of graph methods and active shape and appearance models [11].

The literature [8] based on random graph methods is one of the first approaches to use graph matching in facial landmark localization. The algorithm treats facial feature landmarks as a random graph and uses a rigorous probabilistic model to score potential matches. Dealing with Gabor jets, Elastic Bunch Graph Matching (EBGM) [12] presented by Wiskott *et al.* is a remarkable study in graph fitting. This method uses graphs with nodes at facial landmarks as the model and each node contains a set of Gabor wavelet coefficients.

The most important representatives of active shape and appearance model methods are Active Shape Model (ASM), Active Appearance Model (AAM) and their kinds of descendants. Xiong [13] proposes a method of building a scatter data interpolation model from key points to obtain the initialized shape and this method defines a 3D general shape to align face shapes. The approach combining ASM with Local Binary Patterns (LBP) is presented by J.K *et al.* in [14] and this method makes ASM more robust to illumination because LBP is a local texture feature descriptor with a good performance to illumination change. However, these improvements are mostly for frontal face, rather than faces with pose, illumination and other challenges.

More recently, some other model-based methods [15-17] have shown promising performances in facial landmark localization. Besides, a variety of machine learning techniques [18], like Support Vector Machine (SVM) and Random Forests, are widely used in facial landmark localization.

In our previous work [19], a rotation factor (R) is presented to initialize the test face. However, there are still bad initialized shapes because we cannot distinguish different poses completely. Meanwhile, it cannot work when there is a big change caused by illumination and expression. In this paper, we present a novel method that is able to accurately localize facial landmarks under the complicated condition, such as pose, expression and illumination variations. Firstly, we distinctively establish the frontal, left-side and right-side model in the training stage, because faces can be divided into frontal, left and right faces. In the search stage, Model Selection Factor (MSF) is utilized to automatically choose the suitable model as the global shape for the face. Namely, the initialized shape of a human face is no longer the average face but the shape relatively matches the face with poses. Secondly, we use Patterns of Oriented Edge Magnitudes (POEM) to replace local texture model of ASM. POEM operator can not only extract texture information from different directions around the landmark, it can also get multi-resolution characteristics by different sizes of cells and circles. It is a robust operator to meet challenges of illumination, pose and facial expression variations. Thirdly, we refine the subtle shape variation as a second localization. A second localization is utilized for some organs and counters to approach the optimal solution. There is no doubt that a second

location provides a more reliable process for accurate localization. Based on these, our proposed method is more applicable than the previous work due to its higher accuracy and outstanding performance in four face datasets.

## 2. Active Shape Model

ASM [20] is an invaluable tool to accurately localize feature points. Two sub models of ASM are global shape model and local texture model. Global shape model aims at making the face in the qualified range and describing the whole face. Local texture model is to describe the texture information of each feature landmark. The detailed descriptions are as follows.

### 2.1 Global shape model

The specific steps are as follows:

1) We label landmark points for the training set. Each face is described by a 2D shape vector  $\mathbf{S}_i = (x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_M)^T$  where  $M$  is the number of landmark points and  $(x_1, y_1), \dots, (x_M, y_M)$  are landmarks. Supposing the training sets compromise  $N$  face shapes, Point Distribution Model (PDM) is exploited to describe the face so that the  $N$  – training set can be expressed as a shape vector:  $\mathbf{\Omega} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N\}$

2) Global shape model is established by Eq. (1). We align each face and apply the Principal Component Analysis (PCA) to the aligned training shapes.

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

where  $\mathbf{S}$  is the final localization face, the mean shape  $\bar{\mathbf{S}}$  is the initialized shape,  $\mathbf{P}$  is the feature vector got by dimension reduction with PCA, and  $\mathbf{b}$  is the shape model parameters. The product of  $\mathbf{P}$  and  $\mathbf{b}$  represents possible changes of the initialized shape. In order to guarantee that the new generated shape  $\mathbf{S}$  is reasonable, the elements of  $\mathbf{b}$  are limited within a certain range.

Considering that the training set is mostly based on the frontal shape, the mean shape of the training set is far from the test face with poses. Subsequently, ASM would fail to work because traditional ASM is very sensitive to initialized shape.

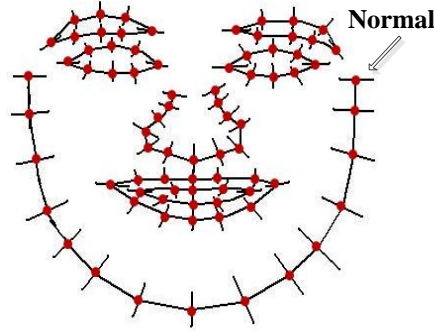
### 2.2 Local texture model

In addition to regard the global shape, local texture model is established for each feature landmark. Local texture model of ASM is the gray-level appearance model that is obtained from pixel profiles.

1) For each landmark, we distinctively choose  $m$  pixels on either side of normal of the landmark. The gray-level profiles  $\mathbf{g}_{i,j}$ , is a  $(2m+1)$ -D vector :

$$\mathbf{g}_{i,j} = [\mathbf{g}_{i,j(1)}, \mathbf{g}_{i,j(2)}, \dots, \mathbf{g}_{i,j(2m+1)}] \quad (2)$$

where  $\mathbf{g}_{i,j}$  is the gray-level profile of the landmark  $j$  in the image  $i$ . Fig. 1 shows the profiles normal to the model boundary.



**Fig. 1.** Profiles normal to the model boundary

2) To reduce the influence of global intensity variations, we compute the derivative  $\mathbf{dg}_{i,j}$  along the profile by **Eq. (3)** and normalize it by **Eq. (4)**:

$$\mathbf{dg}_{i,j} = [\mathbf{g}_{i,j(2)} - \mathbf{g}_{i,j(1)}, \dots, \mathbf{g}_{i,j(2m+1)} - \mathbf{g}_{i,j(2m)}] \quad (3)$$

$$\mathbf{G}_{i,j} = \frac{\mathbf{dg}_{i,j}}{\sum_{k=1}^{2m+1} |\mathbf{dg}_{i,j(k)}|} \quad (4)$$

where  $k$  denotes the  $k^{\text{th}}$  point along the profile of the  $j^{\text{th}}$  landmark in the image  $i$ .

3) The mean normalized derivative profile  $\overline{\mathbf{G}}_j$  is calculated by **Eq. (5)**:

$$\overline{\mathbf{G}}_j = \frac{1}{N} \sum_{i=1}^N \mathbf{G}_{i,j} \quad (5)$$

where  $N$  represents the number of faces.

In the search stage, given a new profile  $\mathbf{G}_j$ , the difference between  $\mathbf{G}_j$  and  $\overline{\mathbf{G}}_j$  can be computed by Mahalanobis distance measure:

$$f(\mathbf{G}_j) = (\mathbf{G}_j - \overline{\mathbf{G}}_j)^T \mathbf{S}_j (\mathbf{G}_j - \overline{\mathbf{G}}_j) \quad (6)$$

where  $\mathbf{S}_j$  is the covariance matrix of the normalized derivative profile of landmark  $j$  for  $N$ -training set.

The value of  $f(\mathbf{G}_j)$  is smaller, the distance to the target landmark is shorter. The shortest distance is regarded as the position of the landmark  $j$ .

Although ASM has a good performance in landmark localization, its localization accuracy is sensitive to initialized shape and factors caused by pose, expressions and illumination variations. The problems mainly appear in two aspects: 1) Initialization. There will be inevitably a local minimum when initialized shape is far from final localization shape. 2) Texture information. The grey-level profiles in traditional ASM is too simple to capture rich texture information of feature landmarks.

### 3. Improved ASM method

To address these problems, we propose a robust ASM method to improve localization accuracy. Firstly, MSF is proposed to automatically select the most suitable model, achieving a reliable initialization for the face. Then, POEM is used to replace local texture model of ASM so that we can achieve the best position of each landmark. Thirdly, a second localization is presented to discriminatively refine the subtle shape variation for some organs and contours.

#### 3.1 Robust initialization via Model Selection Factor

Initialization, which promotes the performance and prevents fitting process from falling into local minima, is the first and key step in landmark localization. It is well known that localization accuracy of ASM is heavily depends on the initialized shape. However, the initialized shape achieved by ASM is conventionally frontal because most of the training sets are frontal. However, in practical situation, faces are always with pose variations. That is to say, the initialized shape is far from the true shape. Hence, we need to improve localization accuracy of final shape by achieving an optimal initialized shape.

As we all know, poses can be divided to left, frontal and right faces. Therefore, we train the frontal, left-side and right-side model to discriminatively localize faces with different poses in the training phase. In the process of search, MSF is utilized to automatically choose the suitable model as the global shape for the face. Subsequently, an optimal initialized shape is obtained so that the initialization problem of traditional ASM can be solved efficiently.

In Fig. 2, while inputting a face with pose variation, the initialized shape of traditional ASM is still a frontal face. Nevertheless, we can achieve a shape that almost matches the target shape when adding MSF. The details are as follows:

1) We localize eyes with Adaboost [21]. Notably, we separately train two eyes, which can localize eyes more accurately in the process of Adaboost classifier training.

2) Sideburns are scanned by hybrid projection function [22]. Then, we compute the distance ( $l_1$ ) of left eye to left sideburn and the distance ( $l_2$ ) of right eye to right sideburn respectively.

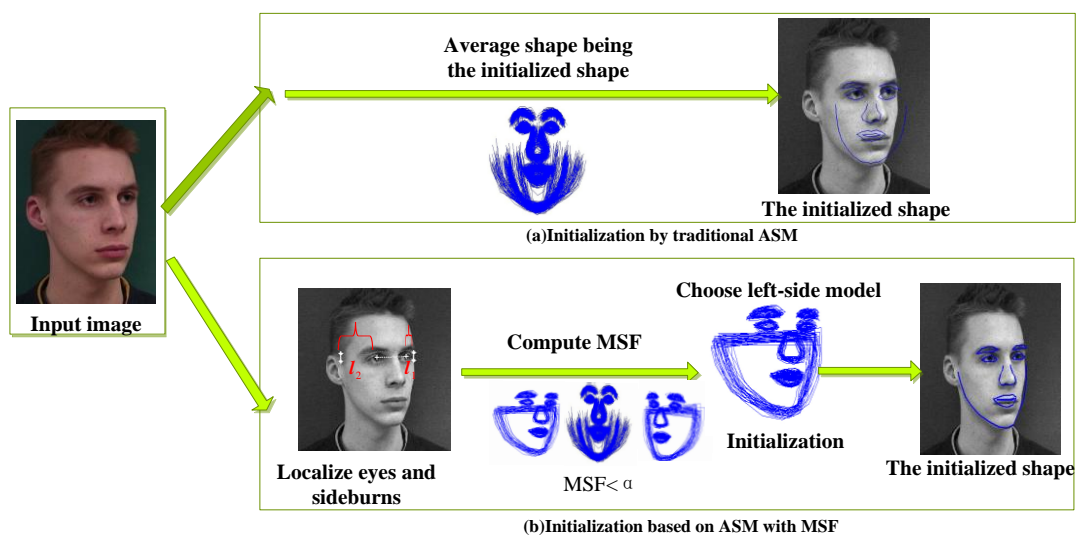
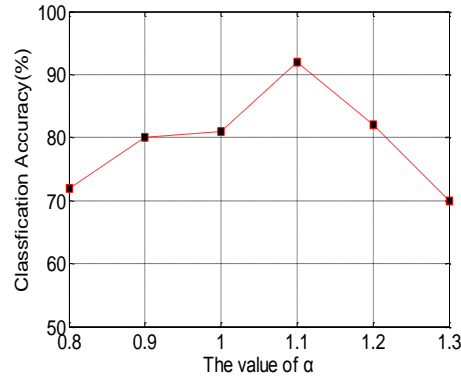


Fig. 2. Initialization process



**Fig. 3.** Relationship between the value of  $\alpha$  and Classification Accuracy

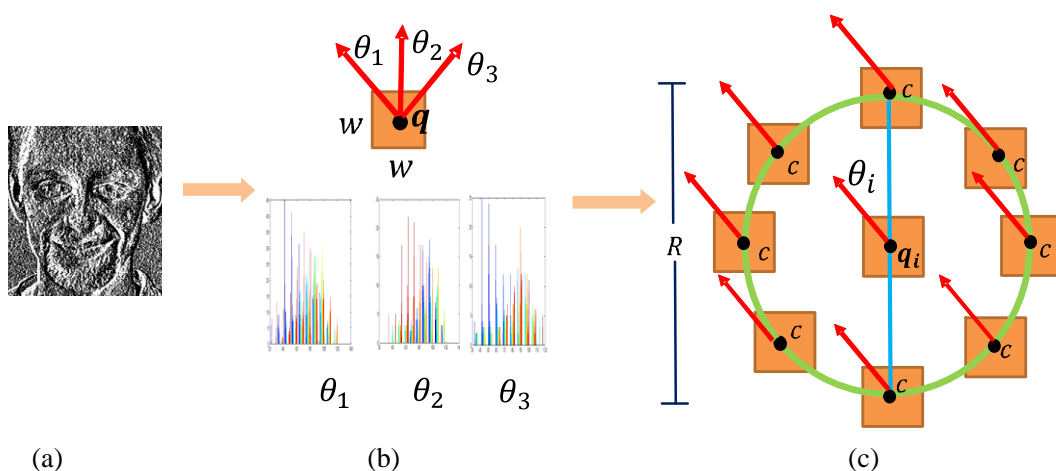
3) MSF is calculated:

$$MSF = \frac{l_1}{l_2} \begin{cases} < \alpha \Rightarrow & \text{left model} \\ = \alpha \Rightarrow & \text{fontal model} \\ > \alpha \Rightarrow & \text{right model} \end{cases} \quad (7)$$

where  $\alpha$  is a threshold whose value is around 1.  $MSF < \alpha$  represents that the face is toward left and left-side model is selected as the global shape model. Meanwhile, the average of all left faces in the training set is used as the initialized shape. Similarly, if  $MSF = \alpha$ , frontal model is selected; If  $MSF > \alpha$ , right-side model is selected. Here,  $\alpha$  is a variable and the values of it are different while applying it to different datasets. For example, **Fig. 3** shows the relationship between the value of  $\alpha$  and Classification Accuracy in IMM dataset. As can be seen in **Fig. 3**, for  $\alpha$  is 1.1, the classification accuracy of MSF can reach 92.1% in IMM dataset. Besides, the values of  $\alpha$  are 0.95, 1.2, 1.15 in CMU PIE, BioID and LFW face datasets respectively.

### 3.2 POEM descriptor for local appearance

Vu *et al.* [23, 24] proposed to apply the LBP-based structure on oriented magnitudes through different orientations to build a novel descriptor called POEM (Patterns of Oriented Edge Magnitudes). It is a robust and fast local texture operator based on the size and direction of texture pixels, which can effectively deal with influences caused by illumination, pose and expression changes. The general process contains two parts: representation of the local details in the cell (**Fig. 4(b)**) and encoding information in the circle with the LBP-based structure. (**Fig. 4(c)**)



**Fig. 4.** The process of POEM feature extraction:(a)gradient image; (b)spatial magnitudes accumulation in a  $w * w$  cell;(c)calculation of oriented magnitudes within a circle

The specific steps are as follows:

1) The gradient magnitudes and orientations of all pixels in the image are calculated. (**Fig. 4(a)**). The gradient orientations  $\theta_i (i=1,2,\dots,m)$  of pixels range from  $0 \sim \pi$ .

2) As seen in **Fig. 4(b)**, we use pixels which are located a  $w * w$  cell centered at pixel  $q$  to build the local histogram. Notably, the histogram calculated within a  $w * w$  cell, is used as the representation of pixel  $q$ .

Precisely, at each gradient orientation  $\theta_i$ , we incorporate gradient information from all cell pixels by computing a local histogram where the contribution of the pixel depends on the gradient magnitude itself. At each pixel of the face, the feature is now a vector of  $m$  values. Here,  $m$  equals to 3. If  $m$  is less than 3, it cannot incorporate sufficient texture information; or  $m$  is larger than 3, it makes POEM sensitive to aging variations.

3) We build the final POEM histogram for each pixel by using the LBP coding process [14] within a circle. Different circles are used to incorporate the accumulated gradient magnitudes over orientations.

For every orientation, the encoding process is shown as **Fig. 4(c)**. The procedures can be depicted by the following equations:

Firstly, at the pixel  $q$ , POEM feature of orientation  $\theta_i$  is calculated by **Eq. (8)**:

$$POEM_{R,w,n}^{\theta_i}(q) = \sum_{j=1}^n f(S(I_q^{\theta_i}, I_{c_j}^{\theta_i}))2^j \quad (8)$$

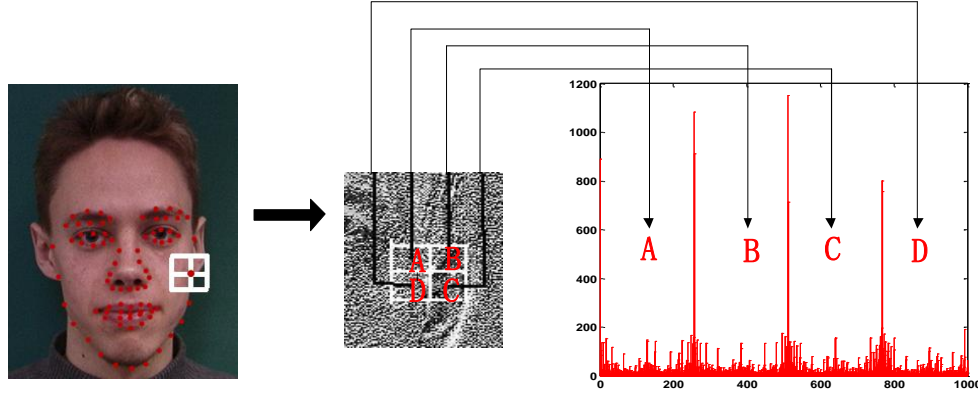
where  $I_q, I_c$  are center pixel and neighborhood pixels respectively and  $n$ , the number of neighborhood pixels, is set to be 8;  $S(.,.)$  is the similarity function measuring the difference of two gradient magnitudes;  $R, w$  refer to the size of circles and cells;  $f()$  is a two-value function based on threshold value  $p$ , defined as:

$$f(x) = \begin{cases} 1, & x \geq p \\ 0, & x < p \end{cases} \quad (9)$$

Then, the final POEM feature of pixel is concatenated to a single histogram sequence through  $m(m=3)$  orientations by **Eq. (10)**:



$$POEM_{R,w,n}(q) = \{POEM_{R,w,n}^{\theta_1}, \dots, POEM_{R,w,n}^{\theta_m}\} \quad (10)$$



**Fig. 5.** ASM with POEM local descriptor

In our work, we apply POEM descriptors to represent local texture model. The number of orientations ( $m$ ) is 3, and the size of circle ( $R$ ) and cell ( $w$ ) is set to 5 and 3 respectively. As shown in **Fig. 5**, we take a  $25 \times 25$  square centered at every landmark firstly. Then, in order to retain spatial information, the square is divided into four regions (A, B, C, D). In each region, we compute the POEM histogram by above steps. Finally, from region A to region D, the POEM histograms are concentrated into a single histogram sequence as the POEM feature for the center landmark. Besides, in the training phase, the mean POEM histogram of landmark  $q$  among  $N$  – training set is calculated by **Eq. (11)**:

$$\overline{POEM}_{R,w,n}(q) = \frac{1}{N} \sum_{i=1}^N POEM_{R,w,n}(q) \quad (11)$$

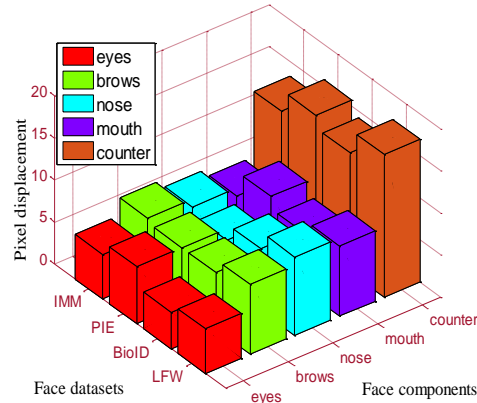
For every landmark, the mean POEM histogram  $\overline{POEM}_{R,w,n}(q)$  among  $N$  – training set is calculated and used as the local representation of that landmark. In the search stage, the similarity between the mean POEM histogram of the landmark and the POEM histograms of the estimated landmark is measured by the Chi square distance [14].

### 3.3 A second localization for subtle shape variation

MSF can achieve a pose-free initialized shape firstly, then we use the global shape model of ASM to describe the whole shape and the texture model POEM to adjust the position of each landmark. We compute mean error of every landmark in each face component. Sixteen landmarks for eyes, sixteen landmarks for brows, twenty landmarks for mouth, thirteen landmarks for nose and fifteen landmarks for face shape counter respectively.

In **Fig. 6**, the z-coordinate means the average pixel displacement between the estimated position and the ground truth, the y-coordinate represents face datasets, the x-coordinate means face components, *e.g.* eyes, eyebrows, nose, etc. It shows the proposed method can achieve accurate localization results in four main face datasets except for the face shape counter. The average error of the counter even reaches to 20 in LFW face datasets. In order to achieve a better performance, the second localization is utilized to improve the localization accuracy for face shape counter. The process of second localization is similar to Section 2.





**Fig. 6.** Localization based ASM with MSF and POEM

## 4. Experimental study of facial landmark localization

### 4.1 Datasets and evaluation metric

To verify our methods, four main face datasets will be introduced in this section: IMM, CMU PIE, BioID and LFW. All of them incorporate different challenges for facial landmark localization. Meanwhile, we compare the proposed method with traditional ASM [20], ASM with LBP [14], ASM+R+POEM [19] and OPM-CDSM [16] (Optimized Part Mixtures and Cascaded Deformable Shape Model).

In our experiment, Mean Average Pixel Error (MAPE) is applied as the error measurement for the facial landmark localization. We define the function as Eq. (11):

$$MAPE = \frac{1}{M} \sum_{i=1}^M \sqrt{(x_p - x'_p)^2 + (y_p - y'_p)^2} \quad (11)$$

where MAPE is the average displacement in pixel between the ground truth position  $(x_p, y_p)$  and estimated position  $(x'_p, y'_p)$ .  $M$  is the landmark number of each image.

Then, some anchor landmarks performances are mentioned because of its importance to the process of localization.

### 4.2 Comparison with Previous Work

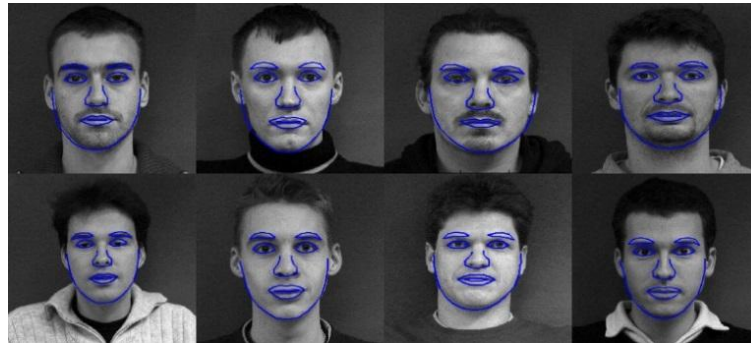
#### Experiment I: IMM face dataset

IMM face dataset, published by Technical University of Denmark, contains 240 images of 40 human faces. Each of which has pose, illumination and facial expression variations.

Firstly, all localization methods will be evaluated for frontal faces. Some frontal faces are shown in Fig. 7 and the localization results on the frontal faces are shown in Fig. 8. It can be seen that the proposed method can achieve an accurate localization. Especially, details such as some organs and counters are accurately captured.



**Fig. 7.** The frontal examples in IMM



**Fig. 8.** Some frontal faces localization results by the proposed method

Secondly, in subsection 4.2, we pick up images under the complicated condition to evaluate all localization methods. Some images are shown in **Fig. 9**. As can be seen from **Fig. 10**, almost every organ and counter are expected to be captured, meaning the proposed method can achieve an optimal localization under complicated condition.



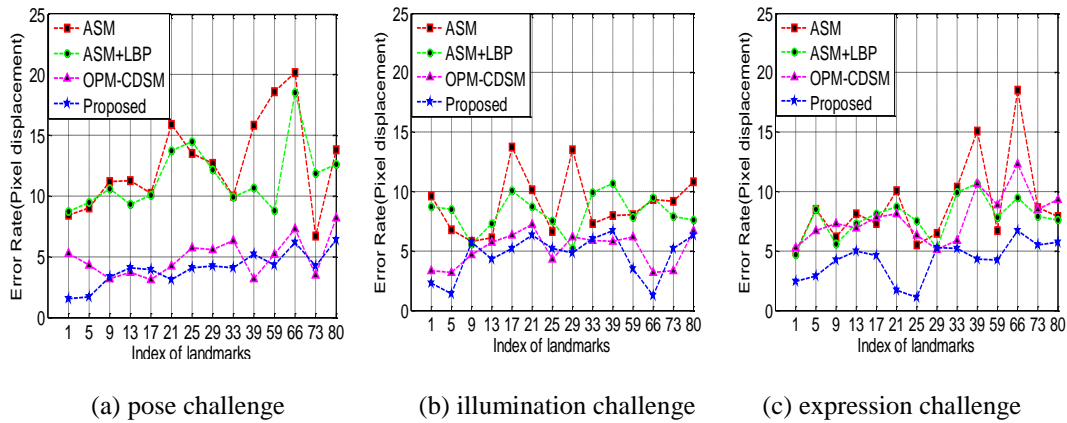
**Fig. 9.** Faces under the complicated condition in IMM



**Fig. 10.** Localization results under the complicated condition in IMM

Finally, we compare the proposed method with traditional ASM, ASM+LBP, ASM+R+POEM and OPM-CDSM during pose, illumination and facial expression changes in a more intuitive way respectively. **Fig. 11** are performances of some anchor landmarks, which play a very important role in localization results. For instance, 1, 5, 9, 13 represents landmarks on eye corners, 17, 21, 25, 29 are landmarks on brows corners, 33, 39 show the mouth corner, 66 and 73 are sideburns, 73 is the chin corner. The y-coordinate means the average pixel displacement between the estimated position and the ground truth. **Fig. 11** show the proposed method achieve a higher localization accuracy in anchor landmarks than other methods.

To verify the efficiency of this facial localization method more sufficiently, we conduct Experiment II, Experiment III and Experiment IV.



**Fig. 11.** Error rates of anchor landmarks



**Fig. 12.** Faces under the complicated condition in CMU PIE

#### **Experiment II: CMU PIE face dataset**

CMU PIE is a dataset of more than 40,000 facial images of 68 people, which includes 13 poses, 43 illumination conditions, and 4 expressions. Some images are shown in [Fig. 12](#).

#### **Experiment III: BioID face dataset**

The dataset consists of 1521 gray level images with a resolution of  $384 \times 286$  pixels. Some images are shown in [Fig. 13](#).



**Fig. 13.** Faces under the complicated condition in BioID

#### **Experiment IV: LFW face dataset**

Labeled Faces in the Wild is a database of face photographs designed for studying the problem of unconstrained face recognition. The dataset contains more than 13,000 images of faces. This is almost the most difficult dataset to localize face because this is outdoor and filled with complex condition together. Some images are shown in [Fig. 14](#).





**Fig. 14.** Faces under the complicated condition in LFW

Firstly, some localization results in Experiment II, Experiment III and Experiment IV are shown in **Fig. 15**. The proposed method also performs well under the complicated condition in the other three face datasets, where faces are in the different sizes, under different poses and illumination variations.



**Fig. 15.** Localization results under the complicated condition in three other datasets

Secondly, from the view of absolute pixel point, we use mean average pixel error (MAPE) to evaluate all methods. As shown in **Table 1**, our method achieves 7.12 on IMM database, 8.0 on CMU PIE, 7.6 on BioID and 7.5 on LFW in terms of MAPE. Although OPM-CDSM leads accuracy in LFW, our methods still control the error in a very low level that consistently outperforms other methods.

**Table. 1** Mean Average Pixel Error (MAPE) on four datasets

Method/ Face datasets	IMM	CMU PIE	BioID	LFW
ASM	12.15	13.7	12.7	10.1
ASM+LBP	10.65	11.5	10.5	9.8
ASM+R+POEM	8.91	10.6	9.7	9.6
OPM-CDSM	8.16	8.3	8.2	<b>4.2</b>
Proposed	<b>7.12</b>	<b>8.0</b>	<b>7.6</b>	7.5

## 5. Summary and conclusions

Traditional ASM as one of the model-based methods, is an efficient landmark localization approach. However, it heavily depends on the initialized shape and is easily influenced by pose, illumination, expression variations. Firstly, in this paper, MSF is presented to automatically select the most suitable global shape model and to achieve a robust initialization. Then, POEM is utilized to replace local texture model of ASM so that we can achieve the best position of each landmark. Finally, we make a second localization for subtle shape variation of some organs and counters, which provides a reliable process for localization. In the experiment, we consider frontal faces and faces with illumination, pose, expression variations in four main face datasets to test. We offer not only the algorithm performance, but also the comparison between the proposed algorithm and other four methods in various respects. The experimental results show that the proposed localization method is robust to illumination, pose and expression challenges. In our future work, the proposed facial landmark localization could be used in some other interesting applications, such as automatic captioning [25, 26] for hearing-impaired users and computational aesthetics [27, 28].

## References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, et al., "Face recognition: a literature survey," *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399-458, 2003. [Article \(CrossRef Link\)](#)
- [2] J. Wang, C. Lu, M. Wang, et al., "Robust face recognition via adaptive sparse representation," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2368-2378, 2014. [Article \(CrossRef Link\)](#)
- [3] M. Wang, B.-B. Ni, X.-S. Hua, et al., "Assistive tagging: A Survey of multimedia tagging with human-computer joint exploration," *ACM Computing Surveys (CSUR)*, vol. 44, no. 4, Article 25, 2012. [Article \(CrossRef Link\)](#)
- [4] M. Wang, R.-C. Hong, X.-T. Yuan, et al., "Movie2Comics: Towards a lively video content presentation," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 858-870, 2012. [Article \(CrossRef Link\)](#)
- [5] J. Yu, Y. Rui, Y.-Y. Tang, et al., "High-order distance-based multiview stochastic learning in image classification," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2431-2442, 2014. [Article \(CrossRef Link\)](#)
- [6] J. Yu, R.-C. Hong, M. Wang, et al., "Image clustering based on sparse patch alignment framework," *Pattern Recognition*, vol. 47, no. 11, pp. 3512-3519, 2014. [Article \(CrossRef Link\)](#)
- [7] J. Yu, Y. Rui, D.-C. Tao, "Click prediction for web image reranking using multimodal sparse coding," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2019-2032, 2014. [Article \(CrossRef Link\)](#)

- [8] T. K. Leung, M. C. Burl, P. Perona, "Finding faces in cluttered scenes using random labeled graph matching," in *Proc. of the fifth IEEE International Conference on Computer Vision*, pp. 637-644, 1995. [Article \(CrossRef Link\)](#)
- [9] D. Cristinacce, T. F. Cootes, "Facial feature detection using AdaBoost with shape constraints," in *BMVC*, pp. 1-10, 2003. [Article \(CrossRef Link\)](#)
- [10] P. N. Belhumeur, D. W. Jacobs, D. Kriegman, et al., "Localizing parts of faces using a consensus of exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2930-2940, 2013. [Article \(CrossRef Link\)](#)
- [11] O. Çeliktutan, S. Ulukaya, B. Sankur, "A comparative study of face landmarking techniques," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 13, 2013. [Article \(CrossRef Link\)](#)
- [12] L. Wiskott, J. M. Fellous, N. Kuiger, et al., "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997. [Article \(CrossRef Link\)](#)
- [13] P.-F. Xiong, L. Huang, C.-P. Liu, "Initialization and pose alignment in active shape model," in *Proc. of Twentieth IEEE International Conference on Pattern Recognition (ICPR)*, pp. 3971-3974, 2010. [Article \(CrossRef Link\)](#)
- [14] J. Keomany, S. Marcel, "Active Shape Models using local binary patterns," *RR 06-07, IDIAP Research institute*, 2006. [Article \(CrossRef Link\)](#)
- [15] X. Zhu, D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2879-2886, 2012. [Article \(CrossRef Link\)](#)
- [16] X. Yu, J.-Z. Huang, S.-T. Zhang, et al., "Pose-Free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (ICCV)*, pp. 1944-1951, 2013. [Article \(CrossRef Link\)](#)
- [17] T. F. Cootes, M. C. Ionita, C. Lindner, et al., "Robust and accurate shape model fitting using random forest regression voting," in *Proc. of ECCV conference*, pp. 278-291, 2012. [Article \(CrossRef Link\)](#)
- [18] J. Yu, D. Tao, "Modern machine learning techniques and their applications in cartoon animation research," *John Wiley & Sons*, 2013. [Article \(CrossRef Link\)](#)
- [19] L.-F. Zhou, B. Fang, W.-S. Li, et al., "Facial Feature Localization Using Robust Active Shape Model and POEM Descriptors," *Journal of Computers*, vol. 9, no. 3, pp. 717-724, 2014. [Article \(CrossRef Link\)](#)
- [20] T. F. Cootes, C. J. Taylor, D. H. Cooper, et al., "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38-59, 1995. [Article \(CrossRef Link\)](#)
- [21] K.-B. Ge, J. Wen, B. Fang, "Adaboost algorithm based on MB-LBP features with skin color segmentation for face detection," in *Proc. of IEEE International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, pp.40-43, 2011. [Article \(CrossRef Link\)](#)
- [22] M. Montazeri, H. Nezamabadi-pour, "Automatic extraction of eye field from a gray intensity image using intensity filtering and hybrid projection function," in *Proc. of IEEE International Conference on Communications, Computing and Control Applications (CCCA)*, pp. 1-5, 2011. [Article \(CrossRef Link\)](#)
- [23] N. S. Vu, A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Proc. of ECCV conference*, pp. 313-326, 2010. [Article \(CrossRef Link\)](#)
- [24] N. S. Vu, A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1352-1365, 2012. [Article \(CrossRef Link\)](#)
- [25] R.-C. Hong, M. Wang, M.-D. Xu, et al., "Dynamic captioning: video accessibility enhancement for hearing impairment," in *Proc. of the international conference on Multimedia, ACM*, pp. 421-430, 2010. [Article \(CrossRef Link\)](#)



- [26] R.-C. Hong, M. Wang, X.-T. Yuan, et al., "Video accessibility enhancement for hearing-impaired users," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 7, no. 1, Artical. 24, 2011. [Article \(CrossRef Link\)](#)
- [27] S. Liu, J. Feng, Z. Song, et al., "Hi, magic closet, tell me what to wear!" in *Proc. of the 20th ACM international conference on Multimedia, ACM*, pp. 619-628, 2012. [Article \(CrossRef Link\)](#)
- [28] Z.-Z Hu, S. Liu, J.-G Jiang, et al., "PicWords: Render a Picture by Packing Keywords," *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 1156-1164, 2014. [Article \(CrossRef Link\)](#)



**Weisheng Li** graduated from the School of Electronics and Mechanical Engineering at Xidian University in July 1997. He received his M.S. Degree and Ph.D. from the School of Electronics and Mechanical Engineering and School of Computer Science and Technology at Xidian University in July 2000 and July 2004, respectively. Currently he is a Professor at Chongqing University of Posts and Telecommunications. His research focuses on intelligent information processing and pattern recognition.



**Lai Peng** received her B.S. degree in Computer Science and Technology from Chongqing University of Posts and Telecommunications, Chongqing, China. She is currently a graduate student at the Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, China. Her research interests include pattern recognition and machine vision.



**Lifang Zhou** was born in Tianshui, Gansu Province, PR China. She received her M.S. degree and Ph.D. degree from the Chongqing University of Posts and Telecommunications in July 2007 and the Chongqing University in December 2013, respectively. Currently she is an Associate professor of Chongqing University of Posts and Telecommunications. Her research focuses on pattern recognition and machine vision, etc.