

# Explicit Occlusion Detection based Deformable Fitting for Facial Landmark Localization

<sup>1</sup>Xiang Yu, <sup>1</sup>Fei Yang, <sup>2</sup>Junzhou Huang and <sup>1</sup>Dimitris N. Metaxas

<sup>1</sup>Department of Computer Science  
Rutgers University

617 Bowser Road, Piscataway, N.J, USA  
{xiangyu, feiyang, dnm}@cs.rutgers.edu

<sup>2</sup>Department of Computer Science  
University of Texas at Arlington

500 UTA Boulevard, Arlington, T.X, USA  
jzhuang@uta.edu

**Abstract**—This paper addresses the problem of facial landmark localization on partially occluded faces. We propose an explicit occlusion detection based deformable fitting model for occluded landmark localization. Most recent shape registration methods apply landmark local search and attempt to simultaneously minimize both the model error and localization error. However, if part of the shape is occluded, those methods may lead to misalignment. In this paper, we introduce regression based occlusion detection to restrict the occluded landmarks' error propagation from passing to the overall optimization. Assuming the parameter model being Gaussian, we propose a weighted deformable fitting algorithm that iteratively approaches the optima. Experimental results in our synthesized facial occlusion database demonstrate the advantage of our method.

## I. INTRODUCTION

Landmark detection and localization is a key component of many computer vision applications[5], [10]. Facial shape registration aims to match a group of predefined landmarks in a given facial image. There are a number of applications of face alignment, i.e. face tracking[1], face modeling[2], facial expression analysis[3], face animation[4], etc, all of which require good precision of facial landmarks. Yet facial landmark localization is a challenging problem because in real world, human faces vary in a wide range due to different reasons, i.e. age, races, expression, pose, lighting conditions, etc, which increase the complexity of the optimization problem.

Many approaches have been proposed to deal with the problem. One of the early methods is Active Shape Model (ASM)[6]. In ASM, a Point Distribution Model (PDM) controls the shape variance and gradient distributions of a set of landmarks. The central idea is using landmarks with certain neighborhood patches to represent an object. By iteratively finding proper match in the neighborhood, the shape is updated. With similar consideration, plenty of works are developed[12], [13]. Another successful model, Active Appearance Model (AAM)[7] considers the appearance as the sum of mean appearance and certain linear combination of basis of the appearance space. By local patch searching, shape parameters are updated to achieve smallest appearance error from the reference appearance. Also there are various constraints[8], [9], [14] added to further improve the model.

This work was supported by the National Space Biomedical Research Institute through NASA NCC 9-58.



Fig. 1. Faces with occlusion

However, in many scenarios faces may be partially occluded by hands, sunglasses, clothes, etc, as shown in Fig.1. Most of the above methods would fail in such cases. If part of the face is invisible, the local patches inside occlusion can not find proper neighborhood to best match pre-trained templates for alignment. And further such incorrect matches yield distorted shape results. Some improvements are proposed to reduce the effects. Zhou et al.[18] proposed a Bayesian inference solution based on tangent shape approximation. Gu and Kanade[17] proposed a generative model for maximizing a posterior. Saragih et al.[19] introduce a robust error function to control the unseen landmarks' variation. The parametric way is data-sensitive and can just deal with restrained region of occlusions. Fei et.al[21] introduced a sparse shape error term to compensate the occluded landmarks' deviation. Such error term may fit the model well. But it may still distort the whole shape. And it can not tell whether the error term is caused by the occlusion or the visible landmarks' noise.

In this paper, we propose a new method that explicitly estimate the likelihood of occlusion for each landmark. The assumption is that occlusion is just a small portion of the whole facial area. Then we run off-line training with positive and negative sample patches to set up such likelihood. For those landmarks which are detected invisible with high probability, they do not participate in the local patch searching. While in the integral deformable fitting, they do not contribute to the global and local parameters' update. Without the influence of occluded patches, optimizing a shape and model structure constrained likelihood is expected

to improve the approximation. To quantitatively evaluate the proposed method, we built three face datasets with synthesized occlusions which approaches natural occlusion. The experimental results reveals the advantage of our method.

The following content in section 2 is the occlusion detection and deformable fitting algorithms. Section 3 illustrates experimental results comparing with a state-of-the-art method. Section 4 is the conclusion of our work.

## II. OCCLUSION DETECTION AND DEFORMABLE FITTING

Assuming a shape defined with  $n$  landmarks, we define the shape vector  $S$  by concatenating the  $x$  and  $y$  coordinates of all the landmarks. For simplicity, we take  $x$  to represent its two dimensional coordinates.

$$S = [x_1, x_2, \dots, x_n] \quad (1)$$

Proposed by Coots et al.[6], ASM captures variant face shapes by a mean shape and a linear combination of  $k$  selected shape basis.

$$S = \bar{S} + \Phi u \quad (2)$$

where  $\bar{S}$  is the mean shape vector,  $\Phi = [\phi_1, \phi_2, \dots, \phi_k]$  the  $k$  shape basis matrix,  $u \in \mathbb{R}^k$  the basis coefficient vector.

The general Point Distribution Model (PDM) taking global transformation into consideration. Our deformable fitting model is based on the PDM. In (3),  $s$  is a scaling factor;  $R$  represents rotation matrix;  $t$  is a translation vector; the content in bracket is the ASM model.

$$x_i = sR(\bar{x}_i + \Phi u_i) + t \quad (3)$$

Actually ASM is depicting the variation in shape, which is non-rigid local deformation inside shape itself. While the PDM considers not only such local deformation but also the global affine transformation. Thus, any face in any pose could be translated to a frontal predefined reference face. The problem is to find the parameters.

$$\check{P} = \{s, R, u, t\} \quad (4)$$

### A. Occlusion Detection

Occlusion is that comparing to fully observable facial images, there are regions which is not consistent to the normal facial appearance. The occlusion may be caused by objects, such as books, scarf, etc. Or it may be caused by parts of human body as well, i.e. hands, beards and so on. We assume that the occluded region takes up only a portion of the whole facial area. If most of the parts are occluded, i.e. only a nose is not occluded, it has little significance in facial applications. Since the initial landmarks are not well aligned, the occlusion detection is not accurate in this step. The way we take is firstly run our deformable fitting method without occlusion information to get more accurate landmarks positions. Then we raise the occlusion detection to exclude certain occluded landmarks and run the fitting method repetitively until shape is converged.

We set each landmark with a certain size rectangle neighborhood. Since most of the rectangles are on facial

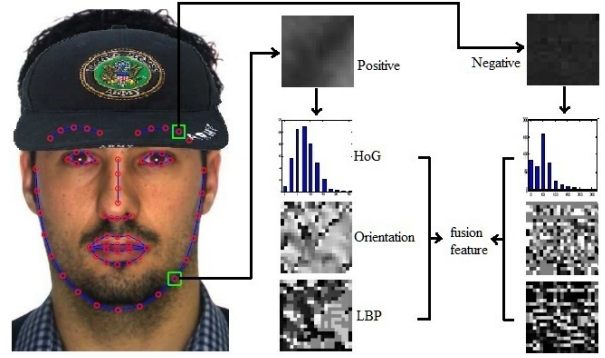


Fig. 2. Facial image positive and negative patches with HoG, Orientation and LBP features

images, the task is becoming discriminating occluded patch appearance from normal facial patches. We notice that some of those facial landmarks have ample information of fixed edges. While occluded patches may have few edge information or the orientation of edges is not consistent to the normal patches. In addition, occlusion texture appears different from the facial texture.

As a result, we adopt Histogram of Gradient (HoG)[16], Orientation[20] and Local Binary Pattern (LBP)[22] features over each landmark patch, which is shown in Fig.2. Then we normalize the three features and concatenate them to form a fusion feature. Such feature combination is expected to capture the magnitude and direction information of the edge pattern and the facial texture pattern.

$$h(X, I) = \sqrt{\Delta I_x^2 + \Delta I_y^2}, X = (x, y) \quad (5)$$

$$r(X, I) = \arctan \frac{\Delta I_y}{\Delta I_x}, X = (x, y) \quad (6)$$

$$b(X, I) = \sum_{X' \in N_8(X)} 2^{\mathbf{I}\{I(X) > I(X')\}} Idx(X'), X = (x, y) \quad (7)$$

In (7), the indicator function  $\mathbf{I}\{\}$  returns 0 if the condition is not satisfied and 1 otherwise.  $N_8(X)$  is the 8-connection neighborhood of  $X$ . Index function  $Idx(X')$  indicates the index of the neighborhood position  $X'$ , which is among 0 to 7. By collecting positive (aligned or misaligned but not occluded) and negative (occluded) samples, we build a logistic regression model in (8).

$$p(o_i = 1 | x_i, I) = \frac{1}{1 + \exp(af + c)}, f = [h, r, b] \quad (8)$$

Here  $p(o_i = 1 | x_i, I)$  is the likelihood of occlusion given  $i$ th landmark position  $x_i$  and facial image  $I$ .  $f$  is the fusion feature of normalized HoG, Orientation and LBP features. Once each landmark likelihood is obtained, we form a diagonal weighting matrix  $W$  for next shape fitting use.

$$W = \text{diag}\{w_1, w_2, \dots, w_n\}, w_i = p(o_i = 0 | x_i, I) \quad (9)$$

## B. Deformable Fitting

From PDM model, our work should further focus on how to optimize the parameters  $\check{\mathbf{P}}$ . Suppose given  $n$  well-aligned landmarks and a facial image. If some of the landmarks are occluded, we are not expecting the occluded nodes to be well-aligned. But the overall shape must be consistently constrained by the PDM. And the alignment event should be modulated by the occlusion event. In other words, if occlusion is detected, the alignment likelihood of current landmark should not contribute to the whole alignment likelihood. We should choose such parameters  $\check{\mathbf{P}}$  to maximize the likelihood in (10).

$$p(\check{\mathbf{P}}|\{v_i = 1\}_{i=1}^n, I) \propto p(\check{\mathbf{P}}) \prod_{i=1}^n [p(v_i = 1|x_i, I)]^{w_i} \quad (10)$$

$v_i$  is a binary discrete random variable indicating whether  $i^{\text{th}}$  landmark is aligned ( $v_i = 1$ ) or misaligned ( $v_i = 0$ ). Each landmark's alignment state is assumed independent from other landmarks' alignment states. Hence, the overall shape's probability of alignment is achieved by multiplying each landmark's alignment likelihood. Here we modulate alignment likelihood by non-occlusion likelihood  $w_i$ . If  $w_i$  is binary distributed, the meaning of (10) is straight forward. When  $w_i$ 's distribution becomes continuous, actually we are modulating the likelihood in a continuous manner, not binary decision any more. Consequently the right part of (10) is derived by Bayesian rule.

The parameters  $\check{\mathbf{P}}$  are set from the PDM model which applies PCA to a set of registered shapes. In common, the distance in PCA subspace is measured by Mahalanobis distance, which is a kernel  $l_2$ -norm measurement. Thus, we assume the prior confirms to Gaussian distribution.

$$p(\check{\mathbf{P}}) \propto \mathcal{N}(\mu; \Lambda), \Lambda = \text{diag}\{\lambda_1, \dots, \lambda_k\} \quad (11)$$

where  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue corresponding to the  $i^{\text{th}}$  shape basis in  $\Phi$  from the nonrigid PCA approach,  $\mu$  is the mean parameter vector respectively.

In (10), eagle eyes may notice the aligned likelihood  $p(v_i = 1|x_i, I)$  is similar to our previous defined occlusion likelihood  $p(o_i = 1|x_i, I)$ . From concept point of view, occluded landmarks have no significance of alignment or misalignment, which means aligned or misaligned ones may not be occluded. Moreover, the alignment likelihood focuses only on whether the landmarks are in proper positions or not. While occlusion likelihood deals with not only the aligned or misaligned (but not occluded) conditions but also types of different occlusion conditions. One may also notice that in occlusion detection training process, we picked not only positive samples but also all kinds of occlusion samples. The more we pick from the occlusion cases, the less likely it is classified as non-occluded. In contrast, the negative samples of misalignment could be any case except the aligned condition.

For alignment likelihood, with similar consideration as the occlusion likelihood, we could also establish logistic regression way to attain the coefficients. The positive samples

could only be aligned patches while negative samples could be any misaligned or occluded case.

$$p(v_i = 1|x_i, I) = \frac{1}{1 + \exp(a'f' + c')} \quad (12)$$

where  $f'$  is fusion of HoG and Orientation feature vector.

Generally given a near-optimal landmark  $x_i$ , we would search its neighborhood to get the optimal alignment likelihood. The possible optimal candidates  $y_i$  form a region  $\Psi_i$ . We assume  $y_i$  confirms to Gaussian distribution with mean  $x_i$  and  $\sigma_i$  standard deviation. Hence, the alignment likelihood can be modeled as a Gaussian Mixture Model (GMM) of the candidates  $y_i$ .

$$\begin{aligned} p(v_i = 1|x_i, I) &= \sum_{y_i \in \Psi_i} p(v_i = 1|y_i, I) p(y_i|x_i, I) \\ &= \sum_{y_i \in \Psi_i} \pi_{y_i} \mathcal{N}(y_i, \sigma_i \mathbf{I}) \end{aligned} \quad (13)$$

where  $\pi_{y_i} = p(v_i = 1|y_i, I)$ . From (13), we rewrite (10) as (14).

$$p(\check{\mathbf{P}}|\{v_i = 1\}_{i=1}^n, I) \propto p(\check{\mathbf{P}}) \prod_{i=1}^n \left[ \sum_{y_i \in \Psi_i} \pi_{y_i} \mathcal{N}(y_i, \sigma_i \mathbf{I}) \right]^{w_i} \quad (14)$$

$$p(v_i = 1, y_i|x_i, I) = p(v_i = 1|x_i, I) p(y_i|v_i = 1, x_i, I) \quad (15)$$

Further we notice in (15)  $p(v_i = 1|x_i, I)$  can be represented as  $\frac{p(v_i = 1, y_i|x_i, I)}{p(y_i|v_i = 1, x_i, I)}$ . An Expectation Maximization (EM) approach is raised to solve the problem. The **E** step is to solve the posterior probability of latent variable  $y_i$  as  $p(y_i|v_i = 1, x_i, I)$ . We denote it as  $\alpha_{y_i}$ .

$$p(y_i|v_i = 1, x_i, I) = \alpha_{y_i} = \frac{\pi_{y_i} \mathcal{N}(y_i, \sigma_i \mathbf{I})}{\sum_{z_i \in \Psi_i} \pi_{z_i} \mathcal{N}(z_i, \sigma_i \mathbf{I})} \quad (16)$$

With intermittent latent variable posterior probability, we approximate  $p(v_i = 1|x_i, I)$  as shown in (17). And the **M** step is to minimize the expectation of negative log likelihood in (18).

$$E_{p(y_i|v_i = 1, x_i, I)}(p(v_i = 1, y_i|x_i, I)) \rightarrow p(v_i = 1|x_i, I) \quad (17)$$

$$\arg \min_{\check{\mathbf{P}}, x_i} E_{q(y)} \left[ -\log \left\{ p(\check{\mathbf{P}}) \prod_{i=1}^n [p(v_i = 1, y_i|x_i, I)]^{w_i} \right\} \right] \quad (18)$$

where  $q(y) = \prod_{i=1}^n p(y_i|v_i = 1, x_i, I)$ . Equivalently, we simplify (18) as (19) shows.

$$\arg \min_{\check{\mathbf{P}}, x_i} \left( \|\check{\mathbf{P}} - \mu\|_{\Lambda^{-1}}^2 + \sum_{i=1}^n w_i \sum_{y_i \in \Psi_i} \frac{\alpha_{y_i}}{\sigma_i^2} \|x_i - y_i\|^2 \right) \quad (19)$$

Here we assume all the landmarks' candidate distribution with the same deviation  $\sigma$ . With property  $\sum_{y_i \in \Psi_i} \alpha_{y_i} = 1$ , The solution of the problem can be formulated as:

$$\Delta \check{\mathbf{P}} = (\sigma^2 \Lambda^{-1} + WJ^T J)^{-1} [WJ^T U - \sigma^2 \Lambda^{-1} (\check{\mathbf{P}} - \mu)] \quad (20)$$

---

**Algorithm 1** Occlusion Detection based Deformable Fitting.

---

- 1: **Input:** initial  $\check{\mathbf{P}}$ , facial image  $I$ .
  - 2: **Output:** optimized  $\check{\mathbf{P}}$ .
  - 3: **Shape Fitting:**  $W = \mathbf{I}$ , run Deformable Fitting (16) and (20)
  - 4:  $\check{\mathbf{P}} \leftarrow \check{\mathbf{P}} + \Delta\check{\mathbf{P}}$
  - 5: **repeat**
  - 6:   **Occlusion Detection:** run (8) and (9) to obtain  $W$
  - 7:   **Shape Fitting:** run Deformable Fitting (16) and (20)
  - 8:    $\check{\mathbf{P}} \leftarrow \check{\mathbf{P}} + \Delta\check{\mathbf{P}}$
  - 9: **until**  $\check{\mathbf{P}}$  converges
- 

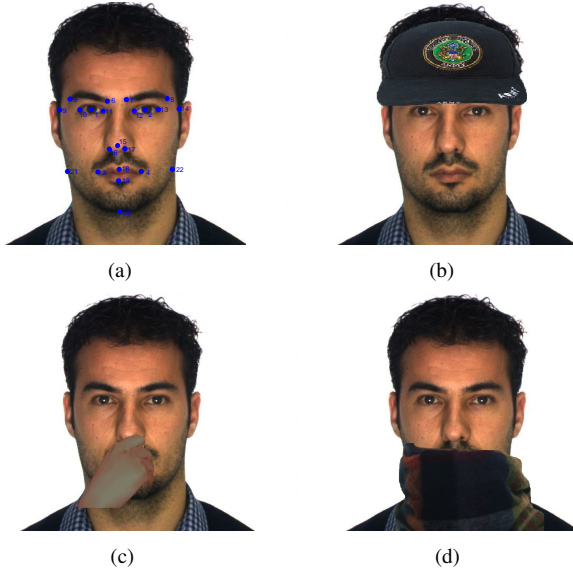


Fig. 3. Faces with ground truth landmarks and artificial occlusions. (a) a face from AR[23] database with 22 blue dot ground truth[11] landmarks. (b) a face with hat occlusion. (c) a face with hand occlusion. (d) a face with scarf occlusion.

where  $J = [J_1, \dots, J_n]$  and  $U = [U_1, \dots, U_n]$ ,  $J_i = \frac{\partial x_i}{\partial \check{\mathbf{P}}}$  and

$$U_i = \left[ \sum_{y_i \in \Psi_i} \frac{\pi_{y_i} \mathcal{N}(y_i, \sigma^2 \mathbf{I})}{\sum_{z_i \in \Psi_i} \pi_{z_i} \mathcal{N}(z_i, \sigma^2 \mathbf{I})} y_i \right] - x_i \quad (21)$$

The whole strategy is summarized in Algorithm.1.

### III. EXPERIMENTAL RESULTS

To evaluate our method, we follow the approach in [21], using face images from AR database[23]. The AR database contains frontal face images of 126 people. Each person has 26 images with different expressions, occlusions and lighting conditions. We select 509 face images from section 1,2,3,5 and use the 22 landmark positions provided by T.F.Cootes[11] as the ground truth. The landmark positions are shown in Fig.3(a).

The occlusion masks are chosen to simulate the occlusions most frequently seen in real world. As shown in Fig.3(b),(c) and (d), there are three types of masks: (1) A hat mask is located just above eyes but occludes all eyebrows;(2) A hand mask is put on mouth which also occludes nose tip and part

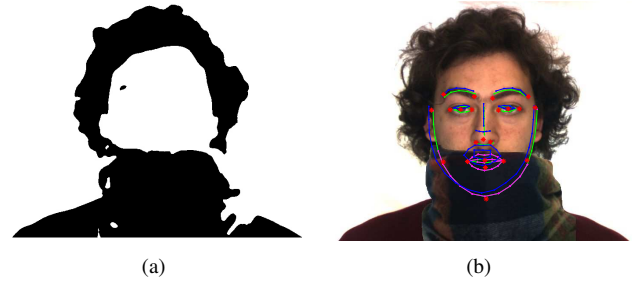


Fig. 4. Visual results of occlusion detection, landmark localization of proposed method and CLM method. (a) result facial image of occlusion detection, black regions indicate occlusions. (b) result facial image of landmark localization, blue dots and lines are of CLM method and green and magenta dots and lines are of proposed method (magenta indicates occluded parts), red stars are the 22 ground truth landmarks.

of chins;(3) A scarf mask lies to occlude the mouth and chin thoroughly. Since we know the ground truth for all occluded landmarks, it is possible to raise quantitative evaluation over the occlusion datasets.

To evaluate landmark precision, we adopt the normalized error metric similar to Jesorsky et al.[26]. The normalized error is the true Euclidean pixel error divided by the distance between eye centers, which is scale invariant. We compare our method with Constrained Local Model (CLM) based method[19], which reports its advantages over ASM[6] and GMM[17].

The visual registration result for one testing image is shown in Fig.4. The ground truth positions are labeled with red stars. The aligned landmarks of CLM is shown in blue lines and the result of our method is shown in green and magenta lines. Green lines show those non-occluded shape registration results and magenta lines indicate occluded landmarks. In the figure, our method correctly discriminate the mouth and the bottom half contour as occluded parts, which appears in magenta. Moreover, the mouth landmarks are in the proper positions comparing with red star points. While the mouth of CLM is pulled up by the edges of nose, which indirectly pushes up the nose landmarks. The CLM implementation is directly from the authors. Since our face model is trained from MultiPIE database[24], there are 66 landmarks as the face shape. Then we create a unique mapping from 66 landmarks to 22 landmarks being consistent with the ground truth labels.

For each of 22 landmarks, we calculate its normalized deviation from true position and accumulate the average error among each of the 3 occlusion datasets. Fig.5 plots such average error for 22 landmarks of 3 occlusion datasets. Blue lines with star dots are proposed result and red lines with circles are CLM's result. The horizontal axis represents index of landmarks. The vertical axis is the normalized error. From the plots, our method is consistently better than CLM in all datasets, which shows the proposed method significantly improves the alignment from one of the state-of-the-art method.

The mean of the average errors for all 22 landmarks is reported in Table.I. We could see for the 3 datasets



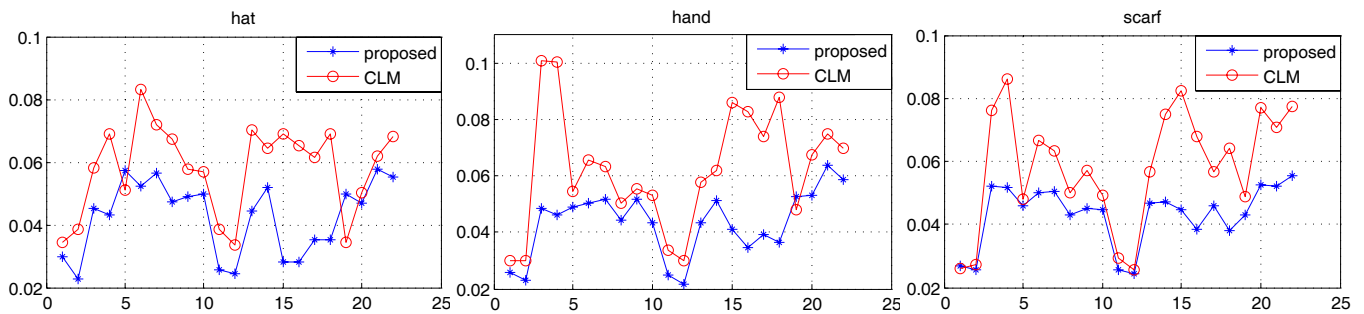


Fig. 5. Normalized average alignment error on 22 landmarks of CLM and our proposed methods on 3 occlusion datasets from AR database[23]



Fig. 6. More visual results of facial landmark localization. The first and second rows are results with hat, hand and scarf occlusion. The third row are results of occluded facial images from LFPW database[25]. Blue dots and lines are results of CLM method. Green and magenta dots and lines are results of proposed method (magenta indicates occluded parts).

and their occluded regions, our proposed method achieves smaller error than CLM’s result. We also notice that for the occluded parts, error is in the same level as the whole dataset’s alignment error with the proposed algorithm. Take hat dataset for example. The average error is  $4.25 \times 10^{-2}$  and the hat occluded area (landmarks 5,6,7,8) error is  $4.74 \times 10^{-2}$ , which increases just  $0.49 \times 10^{-2}$ . While for CLM method, the average error is  $5.79 \times 10^{-2}$ . But the hat occluded region significantly increases to  $8.10 \times 10^{-2}$ . The

same thing happens to the other two datasets, which reflects the effectiveness of our occlusion perception strategy.

Table.II shows average error level for different parts, i.e. contour (landmarks 9,14,20,21,22), eyes (landmarks 10,1,11,12,2,13), mouth (landmarks 3,4,18,19) etc. In hat dataset, occluded part is only eyebrow; in hand dataset, occluded parts are nose, mouth and partial contour (chin); in scarf dataset, occluded parts are nose, mouth and the bottom half of contour. From each part precision point of view, the

TABLE I  
AVERAGE ALIGNMENT ERROR ON 3 DATASETS AND THEIR OCCLUDED PARTS

	Hat	Hat occluded	Scarf	Scarf occluded	Hand	Hand occluded
CLM ( $\times 10^{-2}$ )	5.79	8.10	5.83	7.15	6.27	6.83
Proposed ( $\times 10^{-2}$ )	<b>4.25</b>	<b>4.74</b>	<b>4.31</b>	<b>4.92</b>	<b>4.34</b>	<b>5.33</b>

TABLE II  
FACIAL PART ALIGNMENT ERROR ON 3 OCCLUSION DATASETS

		Contour	Eyebrows	Eyes	Nose	Mouth
hat ( $\times 10^{-2}$ )	CLM	6.03	6.83	4.54	6.51	5.76
	proposed	<b>5.21</b>	<b>5.33</b>	<b>3.28</b>	<b>3.06</b>	<b>4.33</b>
hand ( $\times 10^{-2}$ )	CLM	6.61	5.85	3.92	8.12	8.44
	proposed	<b>5.58</b>	<b>4.89</b>	<b>3.04</b>	<b>3.83</b>	<b>4.60</b>
scarf ( $\times 10^{-2}$ )	CLM	7.15	5.70	3.58	6.90	6.88
	proposed	<b>5.04</b>	<b>4.73</b>	<b>3.23</b>	<b>4.29</b>	<b>4.62</b>

proposed approach also demonstrates its advantage over the CLM method.

Table II further verifies that for those 3 datasets, for CLM method, the occluded parts show more significant error than the non-occluded parts. Take hand database for instance in which nose and mouth are occluded, the nose and mouth's alignment errors are largest. However, the rule is not held for our method. In hand database, the largest error is contour. Nose and mouth's errors keep in the same level as eyebrows and eyes. The reason may be that we detect occlusion ahead and restrict the error of occluded parts from taking into optimizing the whole shape. Moreover, contour and mouth indicate larger misalignment than other facial parts in our method and eyes show the smallest error.

In addition, Fig.6 shows more visual results of the localization. The notations are the same as Fig.4. The first and second rows are the results of the synthesized occlusion facial images from AR database. The third row are the results of challenging occluded facial images from LFPW database[25]. The occlusion detector in natural scene may fail as shown in the third row, especially for the cases with occlusion which has similar appearance as the facial area. Further, as we claimed, our method could deal with cases in which a portion of the face is occluded. If more than half of facial area is occluded, especially the contour and mouth are invisible, the method possibly degrades its performance.

#### IV. CONCLUSIONS

In this paper, we propose a novel occlusion detection based deformable fitting model for occluded facial landmark localization. By explicitly introducing the landmark occlusion detection, our algorithm restricts the occluded landmarks' error propagation from sending to the overall optimization procedure and thus is more robust for landmark localization. Experimental results in the synthesized face occlusion

database demonstrate the advantage of our method. Our future work focuses on facilitating over different occlusions and accelerating the whole pipeline.

#### REFERENCES

- [1] H. Tao and T. Huang. Explanation-based facial motion tracking using a piecewise bezier volume deformation model. *IEEE Comput. Vision and Patt. Recogn.*, CVPR, volume 1, pages 611-617, 1999.
- [2] Medioni, G. Face modeling and recognition in 3-D. *Analysis and Modeling of Faces and Gestures*, AMFG workshop, pages 232-233, 2003.
- [3] Y. I. Tian, T. Kanade and J. F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 23, pages 97-115, 2001.
- [4] Yuenheng Lee, Demetri Terzopoulos and Keith Waters. Realistic modeling for facial animation. *22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH, pages 55-62, 1995.
- [5] T. F. Cootes, D. Cooper, C. J. Taylor and J. Graham. A trainable method of parametric shape description. *Proc. British Machine Vision Conference*, pages 54-61, 1991.
- [6] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38-59, 1995
- [7] T. F. Cootes and G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. the 5th European Conference on Computer Vision*, pages 484-498, 1998
- [8] T. F. Cootes, K. Walker and C. J. Taylor. View-based active appearance models. In *Automatic Face and Gesture Recognition*, pages:227-232, 2000.
- [9] T. F. Cootes, C. J. Taylor. Constrained active appearance models. In *ICCV*, vol. I pages:748-754, 2001.
- [10] A. Blake, M. Isard. *Active Contours*. Springer, 1998
- [11] T. F. Cootes. [http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/tarfd\\_markup/tarfd\\_markup.html](http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/tarfd_markup/tarfd_markup.html).
- [12] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *Proc. the 10th European Conference on Computer Vision*, pages 504-513, 2008
- [13] S. Romdhani, S. Gong, A. Psarrou and R. Psarrou. A multi-view nonlinear active shape model using kernel pca. In *Proc. British Machine Vision Conference*, pages 483-492, 1999.
- [14] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris N. Metaxas and Xiang Sean Zhou. Sparse shape composition: A new framework for shape prior modeling, pages: 1025-1032, CVPR 2011.
- [15] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, 2006.
- [16] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR* 2005.
- [17] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *Proc. the 10th ECCV*, pages I: 413-426, 2008.
- [18] Y. Zhou, L. Gu and H. Zhang. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In *CVPR*, pages I: 109-116, 2003.
- [19] J. Saragih, S. Lucey and J. Cohn. Deformable model fitting by regularized landmark mean-shift. In *IJCV*, 2010.
- [20] Georgios Tzimiropoulos, Stefanos Zafeiriou and Maja Pantic. Robust and Efficient Parametric Face Alignment. In *ICCV*, 2011.
- [21] Fei Yang, Junzhou Huang and Dimitris Metaxas. Sparse shape registration for occluded facial feature localization. In *Conference on Automatic Face and Gesture Recognition*, 2011.
- [22] Xiaoyu Wang, Tony X. Han, Shuicheng Yan. An HOG-LBP Human Detector with Partial Occlusion Handling. In *ICCV* 2009.
- [23] A. Martinez and R. Benavente. The ar face database. Technical report.
- [24] Gross, R., Matthews, I., Cohn, J. F., Kanade, T. and Baker, S. Multi-PIE. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.
- [25] Peter N. Belhumeur, David W. Jacobs, David J. Kriegman and Neeraj Kumar. Localizing Parts of Faces Using a Consensus of Exemplars. In *CVPR*, 2011.
- [26] O. Jesorsky, K. J. Kirchberg and R. Frischholz. Robust face detection using the hausdorff distance. In *Proc. International Conference Audio- and Video-Based Biometric Person Authentication*, pages 90-95, 2001.