Face Alignment Robust to Occlusion

Anonymous

***Abstract*— In this paper we present a new approach to robustly align facial features to a face image even when the face is partially occluded. Previous methods are vulnerable to partial occlusion of the face, since it is assumed, explicitly or implicitly, that there is no significant occlusion. In order to cope with this difficulty, our approach relies on two schemes: one is explicit multi-modal representation of the response from each of the face feature detectors, and the other is RANSAC hypothesize-and-test search for the correct alignment over subset samplings of those response modes. We evaluated the proposed method on a large number of facial images, occluded and non-occluded. The results demonstrated that the alignment is accurate and stable over a wide range of degrees and variations of occlusion.**

I. INTRODUCTION

In this paper we present a new approach to robustly align facial features to a face image even when the face is partially occluded. Previous face alignment methods[1], [2], [3], [4], [5] attempt to optimally fit a regularized face-shape model (such as a PCA-regularized model) using all of the observable local features, typically by iterative gradient search around the initial position obtained by a (mostly-independently run) face detector. Although certain safety measures are often combined, such as adjustable weight on different features depending on features strength, these methods are vulnerable to partial occlusion of the face, since it is assumed, explicitly or implicitly, that there is no significant occlusion and that the given initial position is close to the correct position.

Fig. 1 shows examples of face alignments on non-occluded and occluded facial images. Fig. 1(a) shows face detection results by Haar-like feature based face detector[6]. Fig. 1(b) shows alignment results by BTSM (Bayesian Tangent Shape Model)[2]. As shown in the examples, most of facial features in the left image of the Fig. 1(b) were aligned correctly but lower parts including mouth and jaw were not aligned. In the right image of the Fig. 1(b), the most of the features were not aligned correctly due to occlusion.

In fact, the difficulty that partial occlusion poses is that of cyclic dilemma; occluded or erroneous features should not participate in alignment, but one cannot tell whether a feature is occluded or not unless the correct alignment is known. In order to cope with this difficulty, our approach relies on two schemes: one is explicit multi-modal representation of the response from each of the face feature detectors, and the other is RANSAC hypothesize-and-test search for the correct alignment over subset samplings of those response modes.

Fig. 2 shows an overview of the proposed method. The red boxes in Fig. 2(a) represent search regions for three feature detectors. Although to set search regions is not necessary

(a) Face detection (b) Face alignment

Fig. 1. Face detections by Haar-like feature based face detector[6], and face alignment results by BTSM [2].

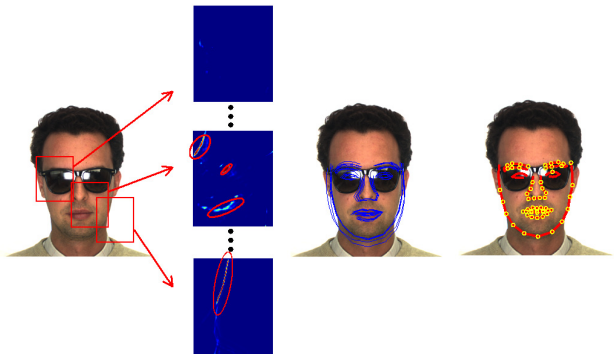
in our method, for the sake of efficiency, the size and the location of the face estimated by the Haar-like feature based face detector[6] is used to set the region on which each feature detector is applied. The alignment process works as follows. Given an image, a large number of detectors, each trained for an individual facial landmark feature, run and produce their response map. In each response map, all the response peaks are identified and localized by the Mean-Shift method, and each peak is described by a 2-D Gaussian model (Fig. 2(b)). The correct alignment we should seek for is the combination of response peaks, one selected from each of visible facial features and none from occluded features, whose locations match well with the face shape model. Since we don't know which features are visible, we employ the RANSAC (Random sampling concensus) strategy. A subset of features is randomly chosen as assumed to be visible, a hypothesis of face alignment is generated from them, and its agreement with other remaining features is tested in terms of the median (rather than mean) of mismatch degrees (Fig. 2(c)). This is repeated until an alignment showing less than acceptable degree of mismatch is found. The alignment thus founds is slightly adjusted by using only the responses of those features whose degree of mismatch is less than the median (that is, those that are identified as visible). Then the shape is refined with additional inliers which are also estimated as visible (Fig. 2(d)).

We evaluated the proposed method on a large number of facial images, occluded and non-occluded. The results demonstrated that the alignment is accurate and stable over a wide range of degrees and variations of occlusion.

To avoid confusion, we defined terminologies as follows: *Feature ID* is a name of a point in PDM (Point Distribution Model) and *Feature point* is an object of a feature ID that the object is a response peak which is descibed by a 2-D Gaussian Model.

II. FACE MODEL

In this paper, non-rigid shape variation of face is represented by the PDM in CLM (Constrained Local Model)



(a) (b) (c) (d)

Fig. 2. An overview of the proposed method. (a) Red boxes represent search regions for three examples of feature detectors. (b) Feature maps of right eye corner (top), a nose part (middle), and cheek part (bottom) detectors. Red circles represent estimated distributions. In the top feature map, any feature points was not detected. (c) Hypotheses of face alignments are shown with blue lines. (d) Red line shows a final alignment after refining. Yellow circles represent feature points which are identified as visible.

framework. The non-rigid shape function can be described as,

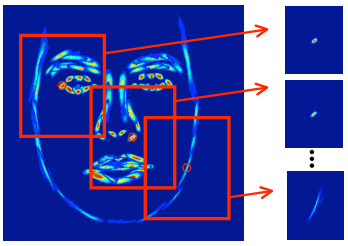
$$\mathcal{S} = \mathbf{X} + \mathbf{\Phi p} \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1^t, \cdots, \mathbf{x}_{N^\mathcal{I}}^t]$, $\mathbf{p}$ is a shape parameter vector, and $\mathbf{\Phi}$ is the matrix of concatenated eigenvectors and the first 4 eigenvectors of $\mathbf{\Phi}$ are forced to correspond to similarity variations. In order to get the $\mathbf{p}$, we employ a CLM as convex quadratic fitting in [1], which is described in the section III.

In the CLM framework, a shape is estimated given feature points from feature detectors. In [1], for the sake of complexity, they used diagonal matrixes to represent feature distributions. However, in terms of locality of feature points, each feature ID has different properties from others which cannot be represented by a diagonal matrix. Fig. 3 shows the distributions of feature responses for all feature IDs. For instance, the distributions of feature IDs in the nostril and eye corner (top and middle images in Fig. 3(b)) are very different from that of feature ID in the cheek region (bottom image in Fig. 3(b)). The eye corner and nostril can be localized accurately since they have clear peaks in the feature responses. However, the feature ID in the cheek region cannot be localized accurately since they have rather long elliptical distribution. Therefore, we use full matrixes to represent the properties accurately and used the properties in shape refinement.

A. Learning Local Feature Detectors

The linear SVMs (Support Vector Machines) are employed for local feature detectors because of its computational efficiency[7]. For the learning of feature detectors, local patches were extracted. Before extracting patches, the images in training data were normalized based on ground-truth. Rotations were normalized by GPA (Generalized Procrustes Analysis)[8], and the scales were normalized on the basis of the size of face region which is given by a face detector.



(a) (b)

Fig. 3. Feature distributions obtained from training data: (a) Feature response of all feature IDs. (b) Selected feature distributions: top, middle and bottom show the distributions of left eye corner, right nostril, and left chin.

Positive patch examples were obtained by centering around the ground-truth, and negative examples were obtained by sampling patches shifted away from the ground-truth. The output of feature detector was obtained by fitting a logistic regression function to the score of SVM and the label {*not aligned (-1), aligned (1)* }[1].

B. Learning Property of Feature ID

Distribution of each feature ID has its peculiar property. To learn the property of kth feature ID, feature response maps of local images around ground-truth are obtained in training data, mean of the feature response maps are calculated, and distribution of the mean is described by a covariance matrix as follows:

$$\Sigma_k = \begin{bmatrix} \mu_{20}/\mu_{00} & \mu_{11}/\mu_{00} \\ \mu_{11}/\mu_{00} & \mu_{02}/\mu_{00} \end{bmatrix} \tag{2}$$

where $\mu_{pq} = \sum_x \sum_y (x-\overline{x})^p (y-\overline{y})^q F_k(x,y)$,$F_k(x,y)$ is feature response value at (x,y) and $(\overline{x},\overline{y})$ is a centroid position of feature response. Fig. 3 shows the distributions for all feature IDs. It clearly shows property of each feature ID.

III. ALIGNMENT FACE WITH OCCLUSION

A. Feature Response Extraction

Unlike the conventional alignment methods which search small area based on a given initial shape[1], [2], [3], initial shape is not required in the proposed method. However, for the sake of efficiency, search window for each feature ID was set according to the face location. Within the search windows, feature responses are calculated and candidate feature points are obtained.

To obtain multiple candidate feature points from a multi-modal distribution of feature response, the feature response were partitioned into multiple segment regions by the Mean-Shift segmentation algorithm[9]. Then the distributions in the segmented regions were approximated through convex quadratic functions as follows:

$$\underset{\mathbf{A}_k^l, \mathbf{b}_k^l, c_k^l}{\operatorname{argmin}} \sum_{\Delta\mathbf{x} \in \Psi_k^l} ||E_k\{Y(\mathbf{x}_k^l + \Delta\mathbf{x})\} - \Delta\mathbf{x}^T \mathbf{A}_k^l \Delta\mathbf{x} + 2\mathbf{b}_k^{lT} \Delta\mathbf{x} - c_k^l||^2 \, subject\ to\ \mathbf{A}_k^l > 0 \tag{3}$$

where Ψ_k^l is a lth segment region in kth feature ID, $\mathbf{x}_k^l$ is the centorid of Ψ_k^l, $E_k()$ is the inverted mutch-score function obtained by applying the kth feature detector to the source image Y, $\mathbf{A}_k^l = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, and $\mathbf{b}_k^l = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

B. Shape Hallucination

In this paper, it is assumed that at least half of feature IDs are not occluded. Let the number of candidate feature points for ith feature ID be $N_i^{\mathcal{P}} (1 < i < N^{\mathcal{I}})$ where $N^{\mathcal{I}}$ is number of the feature IDs. Then, among total feature points $\sum_{i=1}^{N^{\mathcal{I}}} N_i^{\mathcal{P}}$, at least $N^{\mathcal{I}}/2$ feature points and at most $N^{\mathcal{I}}$ feature points exist that support an input facial image. Therefore, correct combination of feature points should be found. In order to find correct combination of feature points, RANSAC[10], [11] was employed.

1) Selecting proposal feature points: In using RANSAC strategy, there are two steps of random samplings in our method. First one is to select M feature IDs. The M is a minimum number of feature IDs that is required to hallucinate a face shape. Second one is to select a feature point among multiple feature points for the each selected feature ID. The first step is to propose a set of non-occluded feature IDs and the second is to propose true feature points among multiple feature responses for the feature IDs.

The sampling is iterated until all the samples are inliers, and the maximum number of iterations k in RANSAC can be determined as follows[10]:

$$k = \frac{\log(1-p)}{\log(1-w)} \tag{4}$$

where p is the probability that the RANSAC selects only inliers in k iteration and w is the probability that all M points are inliers. In our case, w is $\prod_{i=0}^{M-1} \left(\frac{N^{\mathcal{I}}/2-i}{N^{\mathcal{I}}-i} \frac{1}{N_j^{\mathcal{P}}} \right)$ where $N_j^{\mathcal{P}}$ is the number of feature points for jth feature ID and is variable according to the feature ID. For example, let $N_j^{\mathcal{P}}$ for all j are 2 and p, M, and $N^{\mathcal{I}}$ be 0.99, 10, and 83. Then, maximum number of iterations k is larger than 0.9×10^7 and that is not feasible. In our experiments, a proper set of feature points could be selected in 400 iterations in many cases, but in some other cases it could not be done even after $2,000$ iterations.

In order to reduce the number of iteration, we propose a method of coarse filtering out non-candidate feature points. The method is using RANSAC based on the concept of the generalized Hough transform. Two feature points which do not belong to same feature IDs are randomly selected and scale and rotation is calculated. Then, a centroid of face shape is estimated from the two ponts and agreement with centroids obtained from other remaining feature points with respect to the scale and the rotation is tested. The details of the proposed filtering method is as follows:

1 Select two random feature points $\mathbf{y}_i, \mathbf{y}_j$ which do not belong to same feature ID.
2 Estimate scale $\mathbb{S}^l$ and rotation $\mathbb{R}^l$ parameters from two feature points $\mathbf{y}_i, \mathbf{y}_j$ in lth iteration.
3 Calculate distributions of estimated centroid of shape for all feature points: $F(\mathbf{x}) = \sum_{k=1}^{T} f(\mathbf{x}|\mathbf{y}_k, \mathbb{S}^l, \mathbb{R}^l)$ where for given $\mathbb{S}^l$ and $\mathbb{R}^l$ the f is a normal distribution function of centroid for feature point $\mathbf{y}_k$ and $T = \sum_{i=1}^{N^{\mathcal{I}}} N_i^{\mathcal{P}}$. $\mathbf{x}$ represents a position in an image.
4 Get maximum value $\mathbb{C}_{\mathbb{S}^l,\mathbb{R}^l}^l = \max F(\mathbf{x})$ and the position $\mathbf{x}_{max}^l = \arg\max_{\mathbf{X}} F(\mathbf{x})$.
5 Repeat 1 to 4 until the number of iterations reaches given number or the $\mathbb{C}_{\mathbb{S}^l,\mathbb{R}^l}^l$ is larger than a given threshold.
6 Get $\mathbb{S}^{\mathrm{L}}$, $\mathbb{R}^{\mathrm{L}}$, and $\mathbf{x}_{max}^{\mathrm{L}}$ where $L = \arg\max_l \mathbb{C}_{\mathbb{S}^l,\mathbb{R}^l}^l$
7 Given $\mathbb{S}^{\mathrm{L}}$ and $\mathbb{R}^{\mathrm{L}}$, calculate Mahalanobis distance between $\mathbf{x}_{max}^L$ and each feature point with distribution $f(\mathbf{x}_{max}^L|\mathbf{y}_k, \mathbb{S}^l, \mathbb{R}^l)$
8 Take out a feature point that has minimum distance
9 Repeat 6-8 until we get a given number of feature IDs, that the number is at least larger than M.

As a result of the coarse filtering, almost of the selected feature points are inliers. In our experiments, we could get proper inlier set within 5 iterations. Thus, the number of iterations for selecting M correct feature points could be reduced from several hundred to less than five.

2) Hallucinating shapes: From the coarsely filtered feature points and feature IDs, M feature IDs are selected and then M feature points are selected which are associated to the selected feature IDs. Then the parameter $\mathbf{p}$ is calculated explicitly as follows:

$$\mathbf{p} = (\mathbf{\Phi}^t \mathbf{A} \mathbf{\Phi})^{-1} \mathbf{\Phi} \mathbf{b} \tag{5}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{sel_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{A}_{sel_M} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_{sel_1} \\ \vdots \\ \mathbf{b}_{sel_M} \end{bmatrix} \tag{6}$$

The $\mathbf{A}_{sel_k}$ and $\mathbf{b}_{sel_k}$ are $\mathbf{A}_{sel_k}^r$ and $\mathbf{b}_{sel_k}^r$ in (3) which are corresponding to the selected rth feature point of the sel_kth feature point where sel_k is a kth element in a set of selected feature points above and $1 \leq r \leq N_{sel_k}^{\mathcal{P}}$.

Advantage of using this method is that it provides a closed form in calculating shape parameter. In [1], they applied this fitting method iteratively. One of the main reasons that the method required iteration is that only one quadratic fitting function (unimodal distribution) was used for each feature ID even in where there could be multiple strong feature responses (multi-modal distributions). Therefore, it was required to reduce search region iteratively until all quadratic fitting functions localize correct positions with smaller search regions for each iteration. However, our method is to fit a feature response map with multiple fitting functions for each feature ID rather than forces to fit with one fitting function. Thus, once we selected feature points, it can be solved by a closed form.

3) Testing Hallucinated Shape: Since we are assuming that at least half of all feature IDs are not occluded, we can always accept $N^{\mathcal{I}}/2$ feature points that the distances between

the positions of the hallucinated feature points and of feature responses are smaller than median of all the residuals.

The residual between a position of a feature ID of the shape, $\mathbf{x}$ and corresponding feature response of the feature ID $\hat{\mathbf{x}}$ can be calculated as follows:

$$R(\mathbf{x}, \hat{\mathbf{x}}) = (\mathbf{x} - \hat{\mathbf{x}})^t [\mathcal{G}(\Sigma_m) + \Sigma_f]^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \quad (7)$$

where Σ_m and Σ_f are covariance matrixes of shape model and feature response. The $\mathcal{G}$ is a geometric transformation function that is estimated from the hallucination shape. $N^{\mathcal{I}}/2$ feature points that the residual $R(\mathbf{x}_i, \hat{\mathbf{x}}_i)$ are smaller than median value of the residuals for all feature points $\mathbf{x}_i$ are accepted as inliers. Using the inliers, the shape is updated.

The shape is tested in terms of the median of the residuals[12]. Repeat the proposal feature points selection, the hallucination and the test until the median is smaller than a given threshold or the number of the iteration reaches a given number. Then, a hallucinated shape which has minimum error results in an aligned shape with only inlier feature points.

C. Shape refinement

In the previous step, at least $N^{\mathcal{I}}/2$ feature points were selected *safely*. However, there can be more feature points which are not occluded. For example, in case of non-occluded facial image, almost of all feature IDs are shown, but only half of the feature points are selected through previous step. Therefore, it has more feature points to be selected as inliers which are visible.

In order to get more inliers, strong feature responses on perpendicular lines to tangent lines at hallucinated points can be chosen[2]. There are two ways of searching strong feature points along the perpendicular lines.

First one is to search strongest feature response along the perpendicular lines as it is used in [2]. This method uses the feature responses only *on* the perpendicular lines and all the feature IDs were considered to refine a shape. However, under the environment that some parts are occluded, we cannot use all of the feature IDs but need to select some of them according to the feature responses. Feature points which have strong feature responses can be accepted as inliers. However, some feature points which are not occluded are not accepted if strong feature response points are slightly off from perpendicular line. To address this problem, larger area should be considered.

Second one is not to see the feature response only *on* the perpendicular line but also to see neighbors according to properties of feature IDs. As aforementioned, each feature ID has different property of feature response distributions. As shown in the Fig. 2(b) and Fig. 3, it is clear that the feature response of the cheek region is distributed along the edge of cheek which is almost linear shape, while that of the nostril is distributed as almost circular shape. In other words, the feature ID of cheek region cannot be localized easily while that of nostril region can be easily. By seeing its neighbor responses, although a feature response on a perpendicular line is weak, that can be accepted as inliers if its neighbors have strong responses.

The strength of the feature response S on along the perpendicular line is defined as follows:

$$S_i(k) = \sum_{\mathbf{x}} \mathcal{N}_i(\mathbf{x}|k, \Sigma)\Gamma(\mathbf{x}) \quad (8)$$

where the $\Gamma(\mathbf{x})$ and the $\mathcal{N}_i(\mathbf{x})$ are a feature response at $\mathbf{x}$ and the normal probability density function. The i and k are a corresponding ith feature IDs to this feature point and a position on the perpendicular line, respectively.

IV. Error Measurement

The properties of feature IDs that was shown in Fig. 3 give us very interesting point of view on how to measure error between aligned and groundtruth shapes. Conventionally, the error was measured by Euclidean distance. This measurement calculates distance without considering the feature properties. Unlike the feature points of eye corners and lip corners, the feature points of cheek regions and middle of eye-brows are very hard to point out exact positions in image. Many of positions which are on the edge can be answers. However, the conventional Euclidean distance can not handle with these answers.

We propose a new metric based on shape property. It is designed to consider geometrical properties of feature IDs as follows:

1) If a feature ID is positioned on a sharp point such as cornes of eyes, the metric is likely to be a Eucledian distance between a target point and associated ground-truth point.
2) If a feature ID is positioned on a middle of line such as cheek, the metric is a distance between the edge line and a target point.

It is defined as follows:

$$D(i, x) = \min(d(\mathbf{y}_{i-1}, \mathbf{v}_x), d(\mathbf{y}_{i+1}, \mathbf{v}_x)) \quad (9)$$

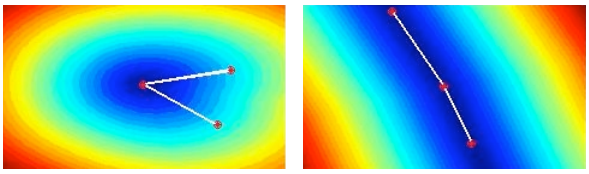
$$d(\mathbf{y}_k, \mathbf{v}_x) = \sqrt{a(\mathbf{y}_k, \mathbf{v}_x)^2 \sin\theta + b(\mathbf{y}_k, \mathbf{v}_x)^2} \quad (10)$$

where

$$a(\mathbf{y}_k, \mathbf{v}_x) = \frac{\mathbf{v}_x \cdot \mathbf{y}_k}{|\ \mathbf{y}_k\ |},$$
$$b(\mathbf{y}_k, \mathbf{v}_x) = |\ c(\mathbf{y}_k, x) - a(\mathbf{y}_k, x)\ |$$
$$c(\mathbf{y}_k, \mathbf{v}_x) = |\ \mathbf{v}_x - \mathbf{y}_k\ |,$$
$$\sin\theta = \sqrt{1 - \left(\frac{\mathbf{y}_k \cdot \mathbf{v}_h}{|\ \mathbf{y}_k\ ||\ \mathbf{v}_h\ |}\right)^2}.$$

The i is an index of ith feature point of ground-truth and $\mathbf{v}_h$ is a tangent line at the ith feature point. The $\mathbf{v}_x$ and $\mathbf{y}_k$ are $\overrightarrow{g_i x}$ and $\overrightarrow{g_i g_k}$, respectively, where the g_k is a ground-truth position of kth feature IDs.

Fig. 4 shows examples of distance metrics for feature IDs on eye corner and cheek. As shown in Fig. 4(b), this measurement allows a point on cheek part to move along the neighbor lines.



(a) (b)

Fig. 4. Distance metrics for feature IDs on eye corner and cheek. The red dots represent positions of previous, current and next feature points in ground-truth.

V. EXPERIMENTS AND ANALYSIS

In order to show robustness of the proposed method, we conducted two experiments in terms of accuracy and stability. The accuracy represents how face alignment is localized accurately and the stability represents how face alignments are stable under occlusions.

For experiment and analysis, two data sets were used: AR database (ARDB)[13], and images which occlusions were generated artificially (AGO). The ARDB includes images of non-occlusion and two types of occlusions: (1) non-occlusion, (2) upper parts occlusion with sun glasses, and (3) lower parts occlusion with scarf. For the AGO images, occlusions were artificially generated from non-occluded face images.

In order to measure error of an alignment shape, we introduce two concepts of errors: (e_n) errors between ground-truth and associated alignment points in non-occluded part and (e_o) errors between those in occluded part. The e_n represents how well the shape is aligned regardless of occlusions. The error e_o represents how well the occluded points are hallucinated. In this paper, the e_n is our main concern.

125 images in the ARDB were used for learning feature detectors and PDM.

A. Accuracy test

In accuracy test, localization accuracy was measured using e_n on the ARDB. Note that the test images includes occluded images as well as non-occluded images.

Fig. 5 shows the result of the accuracy test. The test result using the BTSM[2] is also given for comparison. The x-axis represents images and the y-axis represents mean errors of the alignment shapes. The images in the bottom of the Fig. 5 show face detection (Black line) and the alignment results by the BTSM (Blue line) and the proposed method (Red line). The proposed method (Red circles) outperformed the BTSM (Blue rectangles) for almost of the images regardless of occlusions. As the errors of the alignment by the BTSM are increasing significantly from 27th image, that of the alignment by the proposed method are not increasing. Many of alignment results by the BTSM had large errors due to the initial position problem as well as occlusions.

More examples are shown in Fig. 6. The first row shows face detection results and second and third rows show alignment results by the BTSM and the proposed method. Also, feature points which were identified as visible ones were shown as yellow circles in the third row. The proposed

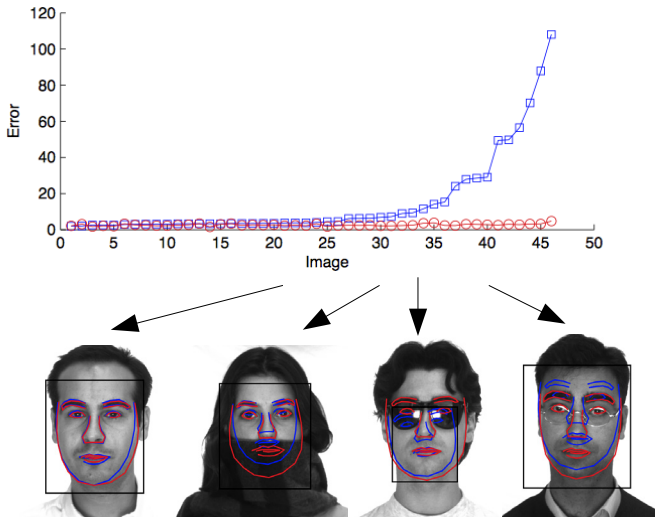


Fig. 5. Result of the accuracy test. The x-axis represents images and the y-axis represents errors of the alignment shapes. x-axis is aligned by the errors of the BTSM results. Blue rectangles and red circles represent the BTSM and the proposed face alignment method, respectively.

method showed very good performance on the occluded images as well as non-occluded images.

B. Stability test

Stability test is to measure robustness of face alignment where various partial occlusions exist. For this test, AGO images were generated as follows. A non-occluded facial image on the ARDB was partitioned manually so that each partition included 5 or 6 facial features in ground-truth. Up to 7 partitions were randomly chosen and the selected partition regions were filled with black so up to half of facial features were occluded. For each degree of occlusion (number of occluded partitions), 5 occlusions of 5 subjects were generated. Some images are show in the bottom of Fig. 7.

In this stability test, the alignment results on an original non-occluded images was used as ground-truth.

Fig. 7 shows test results. The y-axis represents errors measured by (9) and the x-axis represents deree of occlusions. The center line represents the mean errors with maximum and minimum errors. As the degree of occlusions is increasing, the mean error does not chaning a lot. Although the maximum error is increased, that is small. It shows that the proposed method is very stable with respect to the changes of degree of occlusions. The images in the bottom of the Fig. 7 show examples of alignment results on the image with 1, 3, 5, and 7 degrees of occlusions. Feature points which were identified as visible ones were shown as yellow circles.

VI. CONCLUSIONS

In this paper we proposed a new approach of face alignment that is robust to partial occlusion. In the proposed approach, the feature responses from face feature detectors are explicitly represented by multi-modal distributions, hallucinated facial shapes are generated, and a correct alignment is

Fig. 6. Face alignment results on images of ARDB (1-4 columns) and images which occlusions were generated artificially.: The first row shows face detection result. The second and the third rows show alignment results by the BTSM and the proposed methods.

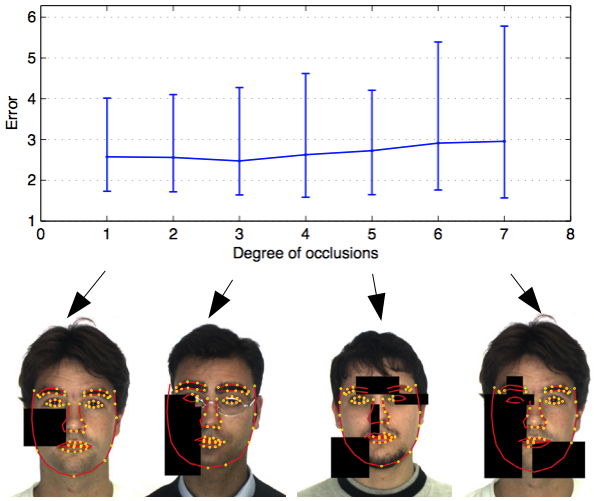


Fig. 7. Result of the stability test. The y-axis represents errors and the x-axis represents deree of occlusions. The center line represents the mean errors with maximum and minimum errors. The images in the bottom show examples of alignment results with showing feature points which were identified as visible ones by yellow circles.

selected by RANSAC hypothesize-and-test. A coarse inlier feature points selection method which reduces the number of iteration in the RANSAC and a new error measurement metric that is based on shape properties are also introduced.

We evaluated the proposed method on ARDB and images which occlusions were generated artificially. The experimental results demonstrated that the alignment is accurate and stable over a wide range of degrees and variations of occlusion.

REFERENCES

[1] Y. Wang, S. Lucey and J. F. Cohn, "Enforcing Convexity for Improved Alignment with Constrained Local Models," *in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[2] Y. Zhou, L. Gu, and H. Zhang, "Bayesian Tangent Shape Model: Estimating Shape and Pose Parameters via Bayesian Inference," *in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, Wisconsin, 2003, pp. 109-116.

[3] T.F. Cootes, C.J. Taylor and Manchester M. Pt, "Statistical Models of Appearance for Computer Vision," 2000.

[4] L. Gu and T. Kanade, "A Generative Shape Regularization Model for Robust Face Alignment," *in Proc. of the 10th European Conference on Computer Vision*, 2008.

[5] D. Cristinacce, T.F. Cootes, "Feature Detection and Tracking with Constrained Local Models", *in British Machine Vision Conference*, 2006, pp. 929-938.

[6] P. Viola, M. Jones, Robust Real-Time Face Detection, *Int. Journal of Computer Vision*, vol. 57, no. 2, 2004, pp. 137-154.

[7] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[8] A. Ross, Procrustes analysis, *Technical Report, Department of Computer Science and Engineering*, University of South Carolina, SC 29208.

[9] D. Comaniciu and P. Meer, Mean Shift: A Robust Approach Toward Feature Space Analysis, *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 24, 2002, pp 603-619.

[10] M.A. Fischler and R.C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. of the ACM*, vol. 24, 1981, pp. 381395.

[11] Y. Li, L. Gu, and T. Kanade, "A Robust Shape Model for Multi-View Car Alignment," *in Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2009.

[12] P.J. Rousseeuw. *Robust regression and outlier detection*, Wiley, New York, 1987.

[13] A.M. Martinez and R. Benavente, "The AR Face Database," CVC Tech. Report #24, 1998.