Joint Face Alignment: Rescue Bad Alignments with Good Ones by Regularized Re-fitting

Xiaowei Zhao[1,2], Xiujuan Chai[1], and Shiguang Shan[1]

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China
mathzxw2002@gmail.com, {chaixiujuan,sgshan}@ict.ac.cn

Abstract. Nowadays, more and more applications need to jointly align a set of facial images from one specific person, which forms the so-called *joint face alignment* problem. To address this problem, in this paper, starting from an initial face alignment results, we propose to enhance the alignments by a fundamentally novel idea: *rescuing the bad alignments with their well-aligned neighbors*. In our method, a discriminative alignment evaluator is well designed to assess the initial face alignments and separate the well-aligned images from the badly-aligned ones. To correct the bad ones, a robust regularized re-fitting algorithm is proposed by exploiting the appearance consistency between the badly-aligned image and its k well-aligned nearest neighbors. Experiments conducted on faces in the wild demonstrate that our method greatly improves the initial face alignment results of an off-the-shelf facial landmark locator. In addition, the effectiveness of our method is validated through comparing with other state-of-the-art methods in joint face alignment under complex conditions.

1 Introduction

In recent years, with the popularity of various digital cameras and the rapid growth of internet-based photo sharing, collecting multiple images of one specific person becomes easier and easier. With the collected multiple facial images for one person, a lot of new applications emerged, e.g., face recognition with image sets [1], face photo album management [2], and person-specific face modeling [3]. These applications appeal for aligning simultaneously the multiple face images of the same person, i.e., the so-called "joint face alignment" problem. Overall speaking, joint face alignment still remains challenging as the appearance of faces can vary a lot due to tremendous variations in pose, expression, occlusion, and low quality imaging, etc.

Joint face alignment problem is initially studied by Learned-Miller's influential congealing procedure, which registers a batch of images by minimizing the entropy of each column of pixels through the image set [4]. Congealing has been proven to work well on simple images, such as binary handwritten digits. Later on, many efforts have been devoted to make congealing more robust to

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part II, LNCS 7573, pp. 616–630, 2012.

complex real-world faces [5,6], which include exploring more robust features for congealing algorithm instead of raw pixels [5] or using least square constrains to estimate warping parameters [6]. Besides these congealing-style methods, Robust Alignment by Sparse and Low-rank Decomposition (RASL) is proposed to make more robust face alignment in the case of occlusions and large lighting changes by minimizing the rank of the image ensemble [7]. However, due to the ignorance of non-rigid transformations, the above methods are restricted from being applied to scenarios which need more accurate face alignment, such as face swapping [8] and face animation [9]. Moreover, the human face information, such as the shape and appearance models of the faces, is not exploited to remove outliers produced in the process of joint face alignment.

Conventional model-based face alignment methods, such as Active Shape Model (ASM) [10] and Active Appearance Model (AAM) [11], are typical non-rigid face alignment techniques, which can also be applied to jointly align a batch of images [12,13]. A straightforward idea is to perform face alignment on each image independently, and jointly align the ensemble of images by registering the facial landmarks of each image with the landmarks of a reference template image. This kind of methods is effective to handle non-rigid pose and shape variations and the incorrect facial landmark configurations can be moderated by exploiting the learned statistical shape or appearance models. However, it is still challenging for these model-based methods to align real-world face images in spite that many efforts have been dedicated to improve the accuracy and efficiency of ASM and AAM (e.g., [14,15,16,17]). In addition, the model-based methods ignore the appearance consistency of the person-specific image ensemble when performing joint face alignment.

More recently, there appeared a few methods that simultaneously consider the human face model and the appearance consistency of image ensemble [18,19]. For example, Zhao et al. perform non-rigid joint face alignment by combining a generic deformable face model (AAM) with RASL [18]. Through the combination, on one hand, RASL is extended from rigid to non-rigid; on the other hand, the generic appearance model (AAM) is regularized to fit adaptively to an unseen face by minimizing the rank of the image ensemble. Tong et al. propose a semi-supervised least-squares congealing algorithm to jointly align an ensemble of facial images [19]. In their method, a small number of images should be manually labeled. During the congealing process, the estimated facial landmark locations are constrained by utilizing an on-line learned non-rigid statistical shape model.

In short, existing joint face alignment methods either depend on an initially labeled image set or regularize the fitting process by minimizing the rank of the whole image set. Different from the above methods, in this paper, we propose a novel method to enhance an initial face alignment results by rescuing the bad alignments with the good ones. Specifically, given a set of facial images from one specific person, we *first* align all the face images in the set with an off-the-shelf model-based facial landmark locator. *Then*, we automatically assess the initial alignments to distinguish the good ones from the bad ones by using

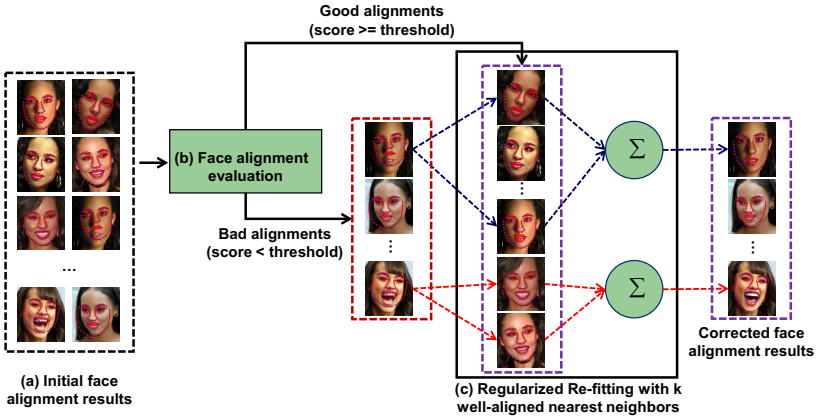


Fig. 1. Overview of the proposed joint face alignment method: (a) initial face alignment results, (b) face alignment evaluation, (c) correcting the bad alignment by regularized re-fitting with k well-aligned nearest neighbors

a simple but robust discriminative face alignment evaluator, which can assign each alignment result a confidence score highly related to the degree of alignment accuracy. *Subsequently*, for each badly-aligned image, we find its k well-aligned nearest neighbors in the image set. *Finally*, through regularizing the shape and appearance space of the badly-aligned image with the selected good neighbors, the facial landmark locations of the badly-aligned image are corrected by our regularized re-fitting algorithm. Experiments conducted on faces in the wild show that our algorithm greatly improves the initial face alignment results. Besides, the effectiveness of our method is also demonstrated through experiments in comparison with competitive methods in joint alignment of complex faces.

Briefly speaking, the main contributions of this paper are: (1) We propose a novel fundamental idea and framework for joint face alignment, i.e., rescuing the bad alignments with their good neighbors. (2) We design a discriminative face alignment evaluator to assess the face alignment. (3) We propose a regularized re-fitting algorithm to correct the bad alignments by the good ones.

The remaining part of this paper is organized as follows. Section 2 gives a brief overview of our joint face alignment method. Section 3 presents more details of our method, including the initialization of face alignments by an off-the-shelf facial landmark locator, the alignment results evaluator, and the regularized re-fitting algorithm. Section 4 reports the experimental results and also the comparisons with the state-of-the-art methods. Section 5 concludes the paper.

2 Method Overview

The basic idea of our method is shown in Fig.1. Simply speaking, our method consists of the following three steps: (1) initialized face alignment, (2) face

alignment results evaluation, (3) correcting the bad alignment by regularized re-fitting with k nearest good neighbors.

In the first step, as shown in Fig.1(a), we first perform face alignment on all the images in the set by using an off-the-shelf facial landmark locator. In this study, Gu's method [16] is exploited for this purpose.

In the next step, as shown in Fig.1(b), a pre-learned face alignment evaluator is exploited to assess the initial alignment results for each image in the set. The well-aligned images, i.e., those with high alignment confidence scores, are automatically selected. For convenience, we denote this image subset as **G**.

Then, in the last step, the images with scores less than a threshold, i.e., badly-aligned, are further processed to obtain enhanced alignment results. As illustrated in Fig.1(c), for each badly-aligned image, we find its k nearest neighbor images from **G**, which are then used to correct the bad alignment by regularizing its shape and appearance space. In this study, a better face alignment result is finally obtained by solving an optimization problem. It is worth noting here that, our good alignment set **G** is dynamically updated. In other words, once the alignment score of an image after rescuing is larger than the threshold, it is added to the set **G**. With this strategy, more good alignment samples can be exploited to rescue the remaining bad alignments.

3 Enhance Joint Face Alignment by Regularized Re-fitting

In this section, we will describe the implementation details of our method.

3.1 Initialization with an Off-the-Shelf Model-Based Face Alignment Algorithm

In this paper, Gu's robust face alignment algorithm is implemented as the off-the-shelf facial landmark locator [16] and used to initialize the face alignment results for the image ensemble. Being one of the state-of-the-art methods, this approach produces impressive results on real world images. The robustness of this method comes from the shape regularization model, which incorporates constrained nonlinear shape prior, geometric transformation and likelihood of multiple candidate landmarks in a three-layered generative model [16].

At the bottom layer of the model, multiple candidate points are provided for each facial landmark, and each candidate is assigned a confidence representing its probability. Let $Q = \{Q_{nk}, n = 1, 2, \cdots, N; k = 1, 2, \cdots, K\}$ denotes the whole candidate set, where N is the number of facial landmarks, and K is the number of candidate points. A binary $N \times K$ assignment matrix h is introduced to assign one candidate position to each landmark. $Q\{h\}$ denote the set of positions selected by h.

The global geometric transformation $\theta = \{R, s, t\}$ is modeled by the middle layer, where a noise term is assigned to each facial landmark to measure its fitting error, and the noise level of each facial landmark point is measured by ρ_n.

The top layer models the prior shape distribution as a mixture of constrained Gaussian [20],

$$p(S|b) = \sum_l \pi_l \mathcal{N}(\Phi_l b_l + \mu_l; \sigma_l^2 I), \tag{1}$$

where the model parameters associated with each Gaussian component are the mixing rate π_l, the linear principal subspace spanned by the columns of Φ_l, the mean shape μ_l, the isotropic shape noise with zero mean and variance $\sigma^2 I$. $b = \{b_l\}$ denotes the deformation coefficients. A latent component variable z_l is also introduced to stands the mixture component label. Λ_l are eigenvalues for the l-th Gaussian component.

Based on the three-layered generative model, the face alignment problem is formulated as a MAP problem, i.e., finding the optimal deformation parameter b and transformation parameter θ by maximizing the posterior $p(b, \theta|Q)$. EM algorithm is used to solve this MAP problem.

In the E-step, taking the expectation $\langle\cdot\rangle$ of the log of $p(b, \theta, S, h, z|Q)$ over the latent variables S, h, z, we can obtain that:

$$\begin{aligned} \Omega(b,\theta) &= \langle \log p(b,\theta,S,h,z|Q)\rangle_{S,h,z} \propto \langle \log p(S|b,z)\rangle \\ &+ \langle \log p(b)\rangle + \langle \log p(z)\rangle + \langle \log p(Q(h)|S,\theta)\rangle + \langle \log p(h)\rangle. \end{aligned} \tag{2}$$

In the M-step, the updating equations for b and θ can be obtained by taking the derivative of (2), i.e.,

$$\frac{\partial \Omega}{\partial b_l} = -(\Lambda_l^{-1} + \frac{\langle z_l\rangle}{\sigma_l^2}) b_l - \frac{\langle z_l\rangle}{\sigma_l^2} \Phi^{\mathrm{T}} (\mu_l - \bar{S}_l) \tag{3}$$

$$\frac{\partial \Omega}{\partial t} = -\sum_n \frac{1}{\rho_n} (-\sum_k Q_{nk} \sum_l \langle z_l\rangle \langle h_{nk}\rangle) - \sum_n \frac{1}{\rho_n} (sR \sum_l \langle z_l\rangle \bar{S}_{nl} + t) \tag{4}$$

$$\frac{\partial \Omega}{\partial s} = -\sum_n \frac{1}{\rho_n} (-\sum_k Q_{nk}^{\mathrm{T}} R \langle h_{nk} S_n\rangle + t^{\mathrm{T}} R \langle S_n\rangle) - s \sum_n \frac{1}{\rho_n} \langle S_n^{\mathrm{T}} S_n\rangle \tag{5}$$

$$\frac{\partial \Omega}{\partial R} = \sum_n \frac{1}{\rho_n} (\langle S_n\rangle t^{\mathrm{T}} - \sum_k \langle h_{nk} S_n\rangle Q_{nk}^{\mathrm{T}}). \tag{6}$$

In the process of face alignment, Zhao's facial landmark detection algorithm [21] is utilized to automatically detect two eye centers and initialize the facial shape.

3.2 Face Alignment Evaluation with a Discriminative Classifier

In this section, a face alignment evaluator is well designed and developed to automatically assess the quality of the initial face alignment results.

In this paper, it is implemented as a boosting-based discriminative classifier, which consists of a global appearance classifier and multiple component-based classifiers, as shown in Fig.2. Specifically, the discriminative classifier is implemented under the cascaded AdaBoost framework [22] and Real AdaBoost [23]

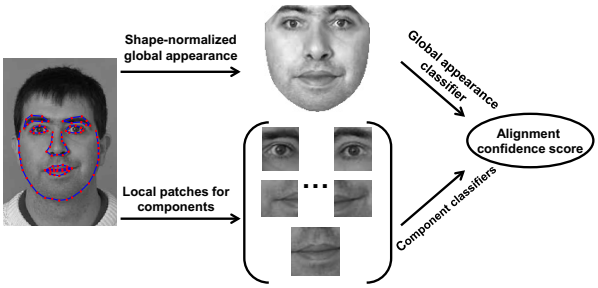


Fig. 2. Illustration of the alignment evaluator, which consists of a global appearance classifier and multiple component-based classifiers

is exploited to assign each alignment result a real-valued confidence score. The alignment confidence score should be correlated with the alignment error. A high score should predict small alignment error and vice visa.

For the global appearance classifier, shape-normalized facial images warped with true alignment are considered as positive class and images warped with perturbed alignment are treated as negative class. Haar-like features are used to characterize the global appearance [22]. This type of global appearance classifier is first used in [14], called Boosted Appearance Model (BAM). However, it shows some limits when being applied to evaluate the face alignment results. As discussed in [15], there is no guarantee that moving along the gradient of BAM's score function will always improve the alignment. It is also verified by our experiments that some alignment results with similar error have very different alignment confidence score, as shown in Section 4.2.

So, in this paper, besides the global appearance classifier, multiple component-based classifiers (e.g., classifiers for eye centers, mouth corners and mouth center) are trained as the compensators to enhance the capability of our face alignment evaluator, as shown in Fig.2. In comparison to the global appearance classifier, the component classifiers characterize more details about the local appearance. For each component classifier, image patches centered at the correct facial landmark location are extracted as positive samples and image patches with several pixels displaced from the true locations are extracted as negative samples.

Given a testing image and its facial landmark locations L, local patches for the components and the shape-normalized global appearance are simultaneously extracted according to L. Then each of them is assigned a confidence value by the corresponding classifier. Finally, the confidence score of L is calculated by the following formula, which is restricted to range $[0, 1]$:

$$f(L) = \frac{1}{1 + \exp\{a[\frac{1}{M}\sum_{i=1}^{M} w_i C_i(L)] + b\}}, \tag{7}$$

where M is the total number of discriminative classifiers, including the global appearance classifier and the component-based classifiers. $C_i(L)$ is the response value of the i-th classifier and w_i is its corresponding weight. Note that in order

to unify the dimension of all the cascaded classifiers, we divide the response value of them by the number of cascaded stages. Moreover, in the following experiments, we set $a = -0.5$, $b = 0$, and $w = 1$ respectively.

3.3 Enhancing Bad Alignment by Regularized Re-fitting with k Nearest Good Neighbors

Based on the initial face alignment results and their alignment confidence scores, a novel regularized re-fitting algorithm is introduced in this section to rescue the bad alignment with its k nearest good neighbors.

Given a testing image I and its k normalized nearest good neighbors $N_i, i = 1, 2, \cdots, k$, the appearance consistency between them is exploited to regularize the appearance space of I , which is formulated as:

$$\min_{\gamma} \sum_{i=1}^{k} \sum_{x} [I(W(x;\gamma)) - N_i(x)]^2, \tag{8}$$

where γ are the parameters representing the shape of image I , and $W(x;\gamma)$ is a non-rigid piecewise affine warping function for the pixel coordinates x.

Therefore, in our solution, the regularization term in (8) is incorporated into the energy function of the generic model-based fitting algorithm and the initial face alignment result of I is corrected by solving the following optimization problem:

$$\max_{\gamma} E(\gamma) - \lambda \sum_{i=1}^{k} \sum_{x} [I(W(x;\gamma)) - N_i(x)]^2, \tag{9}$$

where $E(\gamma)$ is the objective of the generic model-based fitting algorithm and $\lambda > 0$ is the balance factor for the regularization term. Here, the generic model is mainly used to characterize the general shape or appearance variations of face and the regularization term is exploited to adapt the shape or appearance space to image I.

In order to evaluate the effectiveness of the regularization term, we implement our fitting algorithm based on the energy function of Gu's face alignment algorithm and solve it using the generalized EM algorithm [24]. Note that the energy function is *not restricted to* Gu's algorithm but can be any energy term of a model-based face alignment algorithm.

So, replacing $E(\gamma)$ in equation (9) with Gu's energy function, the problem is formulated as follows:

$$\max_{\gamma=b,\theta} \log[p(\gamma, S, h, z|Q)] - \lambda \sum_{i=1}^{k} \sum_{x} [I(W(x;\gamma)) - N_i(x)]^2. \tag{10}$$

In the E-step, taking the expectation of (10) over the latent variables S, h, z, i.e.,

$$\max_{\gamma=b,\theta} \Omega(\gamma) - \lambda \sum_{i=1}^{k} \sum_{x} [I(W(x;\gamma)) - N_i(x)]^2. \tag{11}$$

In the M-step, because problem (11) remains non-linear with respect to variable γ, a common approach to deal with non-linearity is to make first-order approximation and iteratively solves for increments to the parameters $\Delta\gamma$, i.e., maximize the following equation:

$$\max_{\gamma=b,\theta} \Omega(\gamma) + \frac{\partial\Omega}{\partial\gamma}\Delta\gamma - \lambda\sum_{i=1}^{k}\sum_{x}[I(W(x;\gamma)) + \nabla I\frac{\partial W}{\partial\gamma}\Delta\gamma - N_i(x)]^2 \tag{12}$$

with respect to $\Delta\gamma$ and then update $\gamma = \gamma + \Delta\gamma$. Here, $\nabla I\frac{\partial W}{\partial\gamma}$ is the steepest descent image, ∇I is defined as the gradient of I, and $\frac{\partial W}{\partial\gamma}$ is the Jacobian of the warp parameters evaluated at γ.

Thus, the solution of the equation (12) can be computed below:

$$\Delta\gamma = H^{-1}\frac{\partial\Omega}{\partial\gamma} - 2\lambda H^{-1}\sum_{i=1}^{k}\sum_{x}[\nabla I\frac{\partial W}{\partial\gamma}]^{\mathrm{T}}[I(W(x;\gamma)) - N_i(x)], \tag{13}$$

where H denotes the Gaussian-Newton approximation to the Hessian matrix:

$$H = 2\lambda\sum_{i=1}^{k}\sum_{x}[\nabla I\frac{\partial W}{\partial\gamma}]^{\mathrm{T}}[\nabla I\frac{\partial W}{\partial\gamma}], \tag{14}$$

and the expanded form of $\frac{\partial\Omega}{\partial\gamma}$ is given by equations (3), (4), (5), and (6).

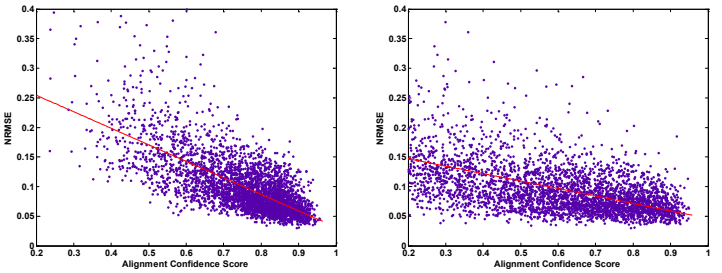
In our implementation, the neighbor samples are determined according to the measurement of the appearance similarity. Here, the simplest Euclidean distance is adopted to characterize the similarity of image intensity. Concretely, the face region is normalized according to positions of the automatically located eye centers. The distance between eye centers is set to 50 pixels.

4 Experiments

In this section, we evaluate the effectiveness of the proposed joint face alignment algorithm through comparing with current state-of-the-art methods. Moreover, the discriminative capability of our face alignment evaluator is also validated in some real world images.

4.1 Data Sets and Evaluation Protocol

In order to quantitatively evaluate the performance of our joint face alignment algorithm and the capability of the discriminative face alignment evaluator, we manually label 3936 images from 20 persons, which are randomly selected from the PubFig [25] database. This database is collected from the internet, which is taken in completely uncontrolled situations with non-cooperative subjects, and covers large variations in pose, expression, lighting and image conditions, etc. For each person, there are about 100∼300 images in PubFig database. In our



(a) Our face alignment evaluator using combined classifiers (b) Face alignment evaluator just using global appearance classifier

Fig. 3. The correlation between the alignment error (NRMSE) and the alignment confidence score. The red line shows the global trend of alignment error with respect to the confidence score.

experiments, 103 facial landmarks are manually labeled along the boundaries of face contour and facial components.

For our algorithm, the normalized root-mean-squared error (NRMSE) relative to the ground truth is adopted as the error measure for the face alignment. The NRMSE is given as a percentage, computed by dividing the root mean squared error by the distance between the two eye centers. The cumulative distribution function (CDF) of NRMSE is used to evaluate the performance of the face alignment algorithm.

4.2 Performance Evaluation of the Face Alignment Evaluator

In this part, we first describe the training process of the face alignment evaluator. Then its performance is examined through quantitative experimental results and visualized examples.

Training for the Face Alignment Evaluator. To train our discriminative face alignment evaluator, about 7000 facial images are collected from various sources, including the CMU PIE database [26], the FRGC database [27], the FERET database [28], and the CAS-PEAL database [29].

For the global appearance classifier, positive samples are generated by warping images with ground truth landmarks and negative samples are generated by perturbing the shape parameters with independent Gaussian noise with variances multiple of the eigenvalues of the corresponding shape bases. Specifically, for each image, about 100 negative samples are generated by randomly adding perturbations to the first seven shape bases. Here, the first four shape bases represent the global scaling transformation, in-plane rotation and translations in horizontal and vertical directions respectively.

In addition, five component classifiers for eye centers, mouth corners and mouth center are trained in our experiments. Besides the image patches centered

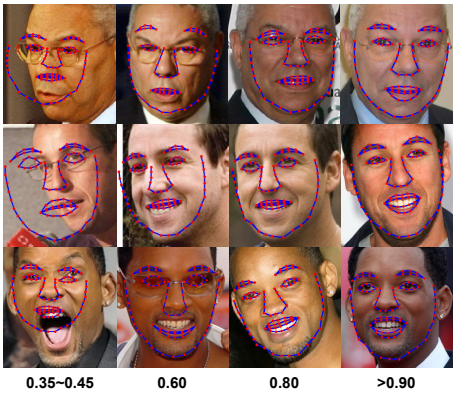


Fig. 4. Example images with the alignment results and the associated confidence scores. Each column shows typical alignment results for the specified score.

at the ground truth position, more synthesized positive samples are generated through transformations, such as shifting by ± 1 pixel, in-plane rotation within 3 degrees. Negative samples are generated by shifting 5$\sim$8 pixels away from the manually labeled ground truth position.

Performance of the Face Alignment Evaluator. We first apply Gu's face alignment algorithm to the randomly selected 3936 images, described in Section 4.1. With the alignment results, the alignment error (i.e., NRMSE) and the alignment confidence score are computed separately. Then, the relationship between these two variables is evaluated quantitatively, as shown in Fig.3. A high score should predict small alignment error and vice visa.

In Fig.3(a), the alignment confidence score is assigned by our face alignment evaluator. It shows that our face alignment evaluator is statistically a good indicator of the alignment accuracy. The alignment confidence score is highly correlated with the alignment accuracy (i.e., the Pearson correlation coefficient between them is -0.6989). However, the distinguishing capability of the face alignment evaluator just using the global appearance classifier is not strong. For example, as shown in Fig.3(b), many alignment results with very similar errors have very different confidence scores and the Pearson correlation coefficient is -0.4929.

More visualized results are shown in Fig.4. Each column gives some representative samples with similar alignment confidence score, which further confirm the validity of our face alignment evaluator in selecting the perfect aligned faces.

Especially, Fig.3 shows that the alignment error is less than 10% of eye-to-eye distance (i.e., the alignment is very close to ground truth annotation) with very high probability when the confidence score is larger than 0.9. So, in the following experiments, we set 0.9 as the threshold to distinguish good alignment results from poor alignment results.

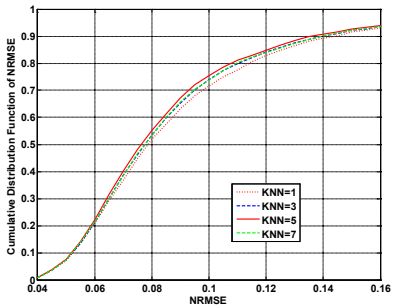


Fig. 5. Performance variation with respect to the number of nearest neighbors

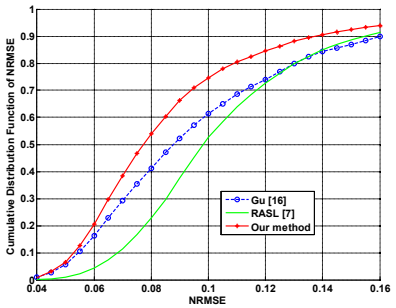


Fig. 6. Comparison with current state-of-the-art methods

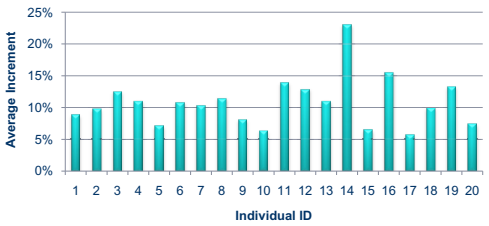


Fig. 7. Alignment accuracy improvement for 20 individuals in comparison with the initial face alignment results

4.3 Performance Evaluation of the Regularized Re-fitting Algorithm

In this section, we evaluate the performance of our method through comparing with Gu's method and RASL, which are state-of-the-art methods for joint face alignment.

It is worth mentioning that, RASL just gives the rigid transformation parameters for each image instead of giving the facial landmark locations. To perform a fair comparison, we first align the image ensemble using RASL. Then, we randomly label five aligned images manually and calculate the mean shape of them. Subsequently, based on the assumption that all aligned images share the same facial landmark configurations, we assign the mean shape to the remaining aligned images. Finally, the facial landmark locations corresponding to the original unaligned images can be calculated by the inverse transformation. Note that, the initial transformation is calculated according to the locations of eye centers.

As for our regularized re-fitting algorithm, the regularization factor λ is set to 1.0. In addition, the alignment performance variation with respect to the number of nearest neighbors is shown in Fig.5. Based on these comparison results, the number of nearest neighbors is set to 5 in the following experiments.

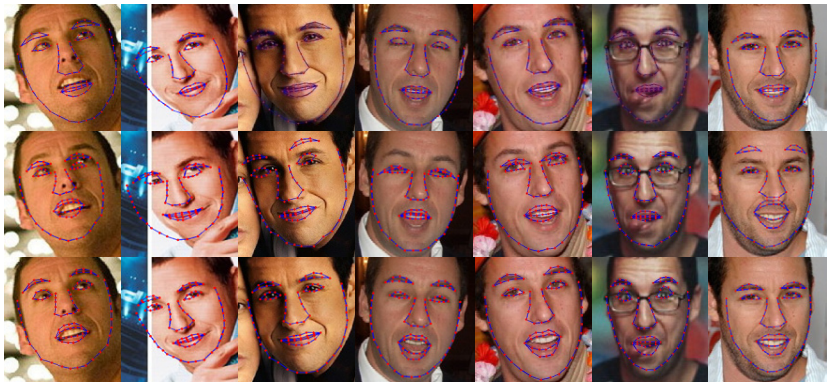


Fig. 8. Alignment results on "Adam Sandler" in comparison with other work. Top row is the alignment results of RASL [7]; second row is the alignment results of Gu's algorithm [16]; the last row is the alignment results of our method.

Our comparison experiments are conducted on the randomly selected 20 image sets, which come from the PubFig database, as described in Section 4.1. The alignment results for these 20 image sets are merged together and the cumulative distribution function of NRMSE is plotted for our method, RASL, and Gu's face alignment algorithm, as shown in Fig.6. It is obvious that our algorithm outperforms the other two kinds of methods significantly. Specifically, our method achieves about 22% higher accuracy than RASL and 13% higher accuracy than Gu's method when the error is less than 10% of eye-to-eye distance.

In addition, the initial face alignment results (i.e., given by Gu's method) for each image set are consistently improved by our regularized re-fitting algorithm. On average, when NRMSE are within $[0.06, 0.14]$, there are about 6%∼23% increment of alignment precision for each of the randomly selected individuals, as shown in Fig.7. The mean and standard deviation of the improvements for all 20 individuals are 11% and 4% respectively.

Some example images from comparisons of our method with Gu's method and RASL are shown in Fig.8. More results of our algorithm on some challenging example images are shown in Fig.9.

5 Conclusion and Future Work

Through enhancing the initial face alignment results of an off-the-shelf model-based facial landmark locator, a promising joint face alignment method is introduced in this paper. To automatically identify images with good face alignment results, an effective discriminative face alignment evaluator is developed in our method. Moreover, a robust regularized re-fitting algorithm is proposed to correct alignment results of the badly aligned images. The principal idea of this paper is to constrain the appearance space of the badly-aligned image by exploring its k nearest well-aligned neighbors and regularize the fitting process of

Fig. 9. More face alignment results of our method on some challenging example images

a general model-based face alignment algorithm. Experiments conducted on the PubFig dataset demonstrate that our method greatly improves the initial face alignment results of an off-the-shelf facial landmark locator and impressively outperforms the state-of-the-art methods in joint alignment of complex faces.

Theoretically, our algorithm may work well as long as there are enough good neighbors to constrain the appearance space of testing images, which is not limited to specific person. In future work, we will further validate this idea through experiments on more general face image set. Moreover, other more robust features (e.g., HOG) will be explored to calculate the similarity between facial images.

Acknowledgments. This work is partially supported by National Basic Research Program of China (973 Program) under contract 2009CB320902; Natural Science Foundation of China under contracts Nos. 61173065, 60833013, and 60832004; and Beijing Natural Science Foundation (New Technologies and Methods in Intelligent Video Surveillance for Public Security) under contract No.4111003.

References

1. Wang, R., Shan, S., Chen, X., Gao, W.: Manifold-manifold distance with application to face recognition based on image set. In: CVPR (2008)
2. Zhu, C., Wen, F., Sun, J.: A rank-order distance based clustering algorithm for face tagging. In: CVPR, pp. 481–488 (2011)
3. Kemelmacher-Shlizerman, I., Seitz, S.M.: Face reconstruction in the wild. In: ICCV, pp. 1746–1753 (2011)
4. Learned-Miller, E.: Data driven image models through continuous joint alignment. IEEE T-PAMI 28, 236–250 (2006)
5. Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: ICCV (2007)

6. Cox, M., Sridharan, S., Lucey, S., Cohn, J.: Least squares congealing for unsupervised alignment of images. In: CVPR (2008)
7. Peng, Y., Ganesh, A., Wright, J., Xu, W.: RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: CVPR, pp. 763–770 (2010)
8. Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P.N., Nayar, S.K.: Face Swapping: Automatically replacing faces in photographs. ACM SIGGRAPH (2008)
9. Saragih, J., Lucey, S., Cohn, J.: Real-time avatar animation from a single image. In: FG (2011)
10. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: Their training and application. CVIU 61, 38–59 (1995)
11. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
12. Cootes, T.F., Marsland, S., Twining, C.J., Smith, K., Taylor, C.J.: Groupwise Diffeomorphic Non-rigid Registration for Automatic Model Building. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 316–327. Springer, Heidelberg (2004)
13. Cootes, T.F., Twining, C.J., Petrovic, V., Schestowitz, R.: Groupwise construction of appearance models using piece-wise affine deformations. In: BMVC, pp. 879–888 (2005)
14. Liu, X.: Generic face alignment using boosted appearance model. In: CVPR, pp. 1079–1088 (2007)
15. Wu, H., Liu, X., Doretto, G.: Face alignment via boosted ranking model. In: CVPR (2008)
16. Gu, L., Kanade, T.: A Generative Shape Regularization Model for Robust Face Alignment. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 413–426. Springer, Heidelberg (2008)
17. Liang, L., Xiao, R., Wen, F., Sun, J.: Face Alignment Via Component-Based Discriminative Search. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 72–85. Springer, Heidelberg (2008)
18. Zhao, C., Cham, W.K., Wang, X.: Joint face alignment with a generic deformable face model. In: CVPR, pp. 561–568 (2011)
19. Tong, Y., Liu, X., Wheeler, F.W., Tu, P.: Automatic facial landmark labeling with minimal supervision. In: CVPR, pp. 2097–2104 (2009)
20. Tipping, M.E., Bishop, C.M.: Mixtures of probabilistic principal component analyzers. Neural Computation 11, 443–482 (1999)
21. Zhao, X., Chai, X., Niu, Z., Heng, C., Shan, S.: Context constrained facial landmark localization based on discontinuous haar-like feature. In: FG, pp. 673–678 (2011)
22. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR, pp. 511–518 (2001)
23. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. The Annals of Statistics 28, 337–407 (2000)
24. Gelman, A., Carlin, J., Stern, H., Rubin, D.: Bayesian data analysis. Chapman Hall, London (1995)
25. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: ICCV, pp. 365–372 (2009)
26. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (PIE) database. In: FG, pp. 46–51 (2002)

27. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: CVPR, pp. 947–954 (2005)
28. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for face-recognition algorithms. IVC 16, 295–306 (1998)
29. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale chinese face database and baseline evaluations. SMC, Part A 38, 149–161 (2008)