

RESEARCH ARTICLE

# A Method for Non-Rigid Face Alignment via Combining Local and Holistic Matching

Yang Yang, Shaoyi Du\*, Zhuo Chen

The School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, ShaanXi, China

\* [dushaoyi@gmail.com](mailto:dushaoyi@gmail.com)



OPEN ACCESS

**Citation:** Yang Y, Du S, Chen Z (2016) A Method for Non-Rigid Face Alignment via Combining Local and Holistic Matching. PLoS ONE 11(8): e0159376. doi:10.1371/journal.pone.0159376

**Editor:** Huiguang He, Chinese Academy of Sciences, CHINA

**Received:** April 28, 2016

**Accepted:** July 3, 2016

**Published:** August 5, 2016

**Copyright:** © 2016 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All training data and part of the testing data are available from the Lucas-Kanade database. Cohn-Kanade database is available in this website: <http://www.consortium.ricmu.edu/ckagree/>. The contact name and email to access data are as following: Nicki Siverling, [kkitani@cs.cmu.edu](mailto:kkitani@cs.cmu.edu), Kanade, T., Cohn, J.F., and Tian, Y.: 'Comprehensive database for facial expression analysis', Automatic Face and Gesture Recognition, 2000, pp. 46-53.

**Funding:** This work was supported by the National Natural Science Foundation of China (61573274), SD; Postdoctoral Science Foundation of China (2015M582661), YY; Natural Science Basic

## Abstract

We propose a method for non-rigid face alignment which only needs a single template, such as using a person's smile face to match his surprise face. First, in order to be robust to outliers caused by complex geometric deformations, a new local feature matching method called K Patch Pairs (K-PP) is proposed. Specifically, inspired by the state-of-art similarity measure used in template matching, K-PP is to find the mutual K nearest neighbors between two images. A weight matrix is then presented to balance the similarity and the number of local matching. Second, we proposed a modified Lucas-Kanade algorithm combined with local matching constraint to solve the non-rigid face alignment, so that a holistic face representation and local features can be jointly modeled in the object function. Both the flexible ability of local matching and the robust ability of holistic fitting are included in our method. Furthermore, we show that the optimization problem can be efficiently solved by the inverse compositional algorithm. Comparison results with conventional methods demonstrate our superiority in terms of both accuracy and robustness.

## 1 Introduction

Face alignment plays an important role in many applications, such as face recognition [1, 2], performance analysis [3] and object tracking. It is an interesting problem and has been paid more attentions in computer vision. The objective of face alignment is to find a transformation between two images, so that all image pixels can be matched to the same semantic locations. In recent years, face alignment has gained significant progress in both theory and practice. However, the problem of only using a single template in face alignment still remains unexplored. The reason could be that the facial appearances exhibit quite differently even if two images are from a same person but with different expressions. In addition, we do not have any prior knowledge about the specific expressions.

The basic face model is one of the most important part in this study. In literature, the well known Active Appearance Model (AAM) [4] solves the problem by synthesizing both an appearance model and a shape model for target face. Many methods [5–8] have been proposed based on the AAM to improve the alignment accuracy and robustness. However, when an unseen target face has a large appearance variation from the training set, this kind of generative appearance model has poor performance [9]. Some discriminative face alignment methods

Research Plan in Shaanxi Province of China (2015JQ6232), YY; and National Natural Science Foundation of China (61379104).

**Competing Interests:** The authors have declared that no competing interests exist.

[10–12] using the regression model directly maps the appearance of an image to the landmarks. They can be successful in face alignment with desired accuracy and robustness as long as there are enough training data. However, they are also critical by the requirement of a large number of training data for regression.

Given only a single template as the appearance information, the current face alignment methods can be roughly classified into two categories: holistic fitting method and local matching method. In the first category, Lucas-Kanade algorithm (LK) [13] is the most popular technique. The matching error under the transformation parameter is directly minimized in the object function. LK can reach sub-pixel accuracy. But the parameter optimization process is based on the gradient descent method which requires good initialization for getting global optimal parameters. Yang et al. [14] proposed a face alignment method using a single template as an extension of the LK. A facial deformation model added as the regularization term is built based on the difference between neutral and expression faces in the linear space. Later Yang et al. [15] also presented a non-linear facial deformation model and a deformation guider in the image alignment formulation. In these methods, the shape parameter is formulated into a more compact space which improve the matching accuracy. But these two methods only effect when the template face is with neutral expression. Zhu et al. [16] proposed a unsupervised face alignment method by nonrigid mapping. A face region is represented by mesh and the mapping between two face images is parameterized by the vertex coordinates. This method is also based on the LK and robust to appearance variations, but it aligns a holistic face area other than each landmark of facial components.

On the other hand, there is a growing focus on local matching methods [17] [18]. This kind of methods use local features to build correspondences between template and target images. The registration process is simplified as calculating the transformation parameters with known correspondence. There are many feature extraction methods such as Scale-Invariant Feature Transform (SIFT) [19], HoG features [20] and Local Binary Features [21]. Wang et al. [22] proposed a patch-based exhaustive search method that could align faces with a single template. Local matching method is more flexible to overcome the limitation of bad initialization. However, since they ignored the global image appearance, the fitting results are dependent too much on the local similarity. Especially, for non-rigid image alignment, the deformation has many degrees of freedom. A large number of local salient features should be found which contain outliers. But, the traditional similarity measure such as Sum of Squared Differences or Normalized Cross-Correlation measure is not robust to outliers caused by nonrigid deformation. A novel similarity measure called Best-Buddies Similarity (BBS) [23] has been proposed recently for its good performance of object detection in the wild. In BBS, the Best-Buddies Pairs (BBP) between two images are found. However, when BBP is used for face images, the matching number decreases under complex geometric deformations, which will influence the precision of non-rigid registration.

In this paper, we propose a method for non-rigid face alignment which only needs a single template, such as using a person's smile face to match his surprise face. Local matching features are introduced into the process of holistic fitting. Both the flexible ability of local matching and the robust ability of holistic fitting are included in our method. We use the basic patch-based local presentation. To avoid negative influence caused by local mismatching, a new local matching search method called K Patch Pairs (K-PP) is proposed. Inspired by the BBP, K-PP is to find the mutual K nearest neighbors between two images. It is robust to outliers caused by complex geometric deformations. Our optimization process is efficient under the Inverse Compositional algorithm. Furthermore, the global rigid transformation is also considered in the alignment framework. Experimental results compared with conventional methods demonstrate the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. In Section 2, we present the problem formulation by a face shape model as the background. In Section 3, we provide the proposed local K Patch-Pair search method. Next, the details about building the local regularization term, parameter optimization and implementation details are given. The method is tested by comparison experiments in Section 4. Finally, conclusions are given in the last section.

## 2 Face Alignment Problem with a Shape Model

Given a template face image  $T$  and a target face image  $I$ , a common way is to formulate the alignment problem in the framework of Lucas-Kanade (LK) algorithm [13]. The goal is to minimize the sum of squared error between two images by finding a warping function  $\mathbf{W}(\mathbf{x}; \mathbf{p})$  with the parameter  $\mathbf{p}$  for all pixel coordinates  $\mathbf{x}$  in the fitting area. The formulation form of the LK algorithm is

$$E_c = \sum_{\mathbf{x}} \| T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \|^2. \tag{1}$$

Let  $\mathbf{s}^p = \{\mathbf{s}_1^p, \mathbf{s}_2^p, \dots, \mathbf{s}_N^p\}$  be the shape feature vector of the  $p$ th face example.  $N$  is the number of annotated face landmarks and  $\mathbf{s}_i = [x_i, y_i]$  recording the geometric location of the  $i$ th landmark. The Point Distribution Model [24] (PDM) is applied on the feature matrix composed by all examples to acquire their mean vector  $\bar{\mathbf{s}}$  and the basis matrix of variation  $\mathbf{Q}$ .

After the model training on face examples, any shape feature  $\mathbf{s}$  can be approximated by the corresponding low-dimensional parameter  $\mathbf{p}$  as:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}\mathbf{p}. \tag{2}$$

With the geometric locations of face landmarks, face region can be separated by computing Delaunay triangulations. In that case,  $\mathbf{W}(\mathbf{x}; \mathbf{p})$  is often chosen as the piecewise affine warp in Eq (1) as the non-rigid transformation.

By optimizing the Eq (1), the face shape can be reconstructed by Eq (2). However, this kind of holistic fitting method is very sensitive to the initial parameters, which is not robust to the non-rigid facial deformation. We are going to add the local matching knowledge to guide the fitting method in the following sections.

## 3 Proposed Method

In this section, We firstly explain our way to search local matching features. Then the optimization algorithm is proposed followed by a description of our formulation model with local constraints. Finally, the novel algorithm and its implementation details are summarized.

### 3.1 Local K Patch Pairs (K-PP) Search

For facial image, the appearance features around some landmarks change a lot during the non-rigid deformation. Local feature matching may have many outliers. Recently, a kind of similarity measure method, called Best-Buddies Pairs (BBP), has made a great process for overcoming the influence of outliers. However, in order to get a high precision of image alignment, two aspects should be considered, 1) the local mismatch rate should be low, and 2) the number of corresponding points should be high. They are two contradictory aspects. Though the BBP can be applied to the object detection successfully, it only ensures the first aspect. This implies that the BBP is not very suit for non-rigid image alignment. To deal with this problem, we propose a K Patch Pairs (K-PP) search method to balance the above aspects by weights.

Taking one landmark as example, finding it's corresponding location by K-PP contains three steps:

1. *Choose patch and search region*

In the template image, we set a  $n \times n$  patch around this landmark. The feature of this landmark  $Pa$  is represented by their RGB values or gray values of all pixels in the patch. We define the search region as neighbor regions  $m \times m$  ( $m \geq 2n$ ) of this landmark. Similarly, in the target image, taking the current or initial landmark as the central, the search region is its  $m \times m$  neighbor regions. We break the search regions both in the template and target images into  $n \times n$  patches by sliding window. Let  $Tpset = \{Tp_1, Tp_2, \dots, Tp_{n_p}\}$  be the patch set in the template search region, and  $Ipset = \{Ip_1, Ip_2, \dots, Ip_{n_p}\}$  be the patch set in the target search region.  $n_p$  is the patch number in search regions.

2. *Find corresponding location*

The conventional BBP method find mutual nearest neighbors (NN) for feature points. In K-PP, we improve the method by finding the mutual  $k$ NN for landmarks patch.

Firstly, the  $k$  nearest neighbors ( $k$ NN)  $\{Ip_j, j = 1, \dots, K\}$  is found in  $Ipset$  for  $Pa$ . Then, for each  $Ip_j$ , if its  $k$ NN in  $Tpset$  includes  $Pa$ ,  $Ip_j$  and  $Pa$  are defined as the mutual  $k$ NN.

Then we find the patch  $Ip_j$  with minimum number  $k$  by Eq (3).

$$\hat{k} = \arg \min_k \{kNN(Tp, Ipset) = Ip_j \wedge kNN(Ip_j, Tpset) = Tp\}. \tag{3}$$

This patch's central is defined as the corresponding location of the landmark.

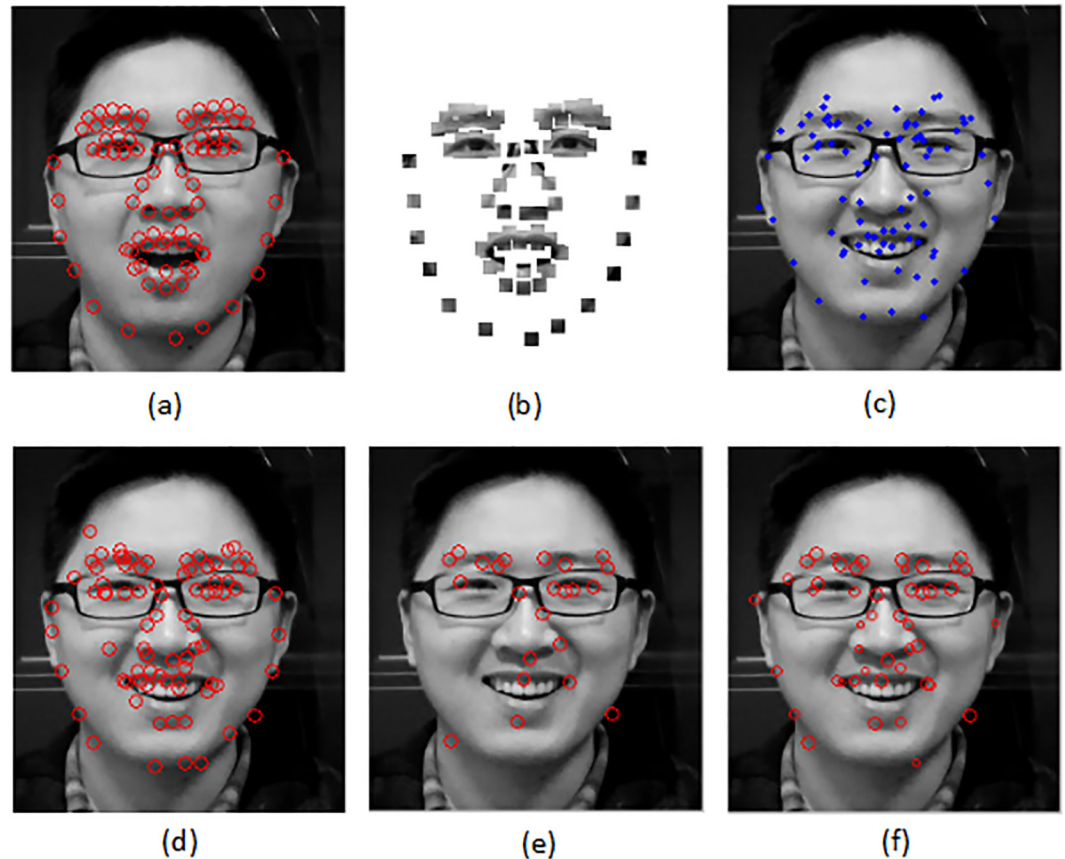
3. *Get weight*

To evaluate the similarity of K-PP, a weight  $w$  is defined based on the parameter  $\hat{k}$ . This weight parameter is useful for the alignment formulation to distinguish the importance of each local corresponding landmarks.

$$w = (K - \hat{k})/K. \tag{4}$$

**Link to the BBP:** Our K-PP is a natural extension of the BBP. But our method is different as follows: (1) The BBP searches the corresponding location for every patch in the template. The K-PP focus on one patch around each landmark. (2) The BBP only considers the mutual NN, which is a particular case of K-PP when  $K = 1$ . (3) The K-PP also give a definition of weight for balancing the similarity and the matching number.

Fig 1 shows the results of local matching search by three different methods. The  $9 \times 9$  patch representation is given in Fig 1(b) around the landmarks. A set of initial landmarks which are far away from the real positions are localized on the target face in Fig 1(c). The local search results are compared in Fig 1(d)–1(f). We can see that the number of corresponding points are the most by the NN method, since it finds the nearest point for every point without other constraints. But there are many mismatching results which will give negative influence to image alignment. The 1-PP which is also the BBP method has a strict rule for choosing matching point to ensure the correct matching. So only a very small number of the landmarks can find the corresponding points. Our K-PP result can get more matching, since we extend the mutual NN in BBP to mutual  $k$ NN. They are not only with high correct rate, but also have the weights to reflect the matching similarity.



**Fig 1. Local matching results.** (a) The template image and landmarks. (b) The patch representation of (a). A  $9 \times 9$  patch is set around each landmark. (c) The target image with initial landmark. (d) The local matching results by NN. (e) The local matching results by 1-PP (KPP,  $K = 1$ ). (f) The local matching results by 5-PP (KPP,  $K = 5$ ). The size of the circle presents the weight value.

doi:10.1371/journal.pone.0159376.g001

### 3.2 Matching Model and Optimization Algorithm

The traditional LK method shown as Eq (1) is easily fallen into a local minimum solution, thus we modified the method by adding a regularization term about the local matching constraint as the prior knowledge. With the local search result by K-PP, we construct the related regularization term. The main purpose here is to minimize the distance from shape features to local search results with weights.

$$E_L = \sum_i w_i (\mathbf{s}_i - \mathbf{s}_{Li})^2, \tag{5}$$

where  $\mathbf{s}_{Li}$  is the feature vector of the  $i$ th landmark of by K-PP search. If we cannot find the K-PP for some feature point, the corresponding vector  $\mathbf{s}_{Li}$  is kept as its initial value, but the weight  $w_i$  is set to be zero.

In order to be presented and calculated easily, let  $R(\mathbf{p}) = \bar{\mathbf{s}} + \mathbf{Q}\mathbf{p} - \mathbf{s}_L$ . Then  $E_L$  can be formulated by matrix.

$$E_L = R(\mathbf{p})^T \mathbf{K}R(\mathbf{p}) \tag{6}$$

where  $\mathbf{K} \in R^{2N \times 2N}$  is the weight matrix according to  $w_i$  defined as a diagonal matrix.

Combined with Eq (1), the proposed formulation method for face alignment with local constraints is

$$E(\mathbf{p}) = E_c(\mathbf{p}) + \lambda E_L(\mathbf{p}), \tag{7}$$

where  $\lambda$  balances between two terms.

Considering the computational efficiency, we adopt the idea of Inverse Compositional(IC) algorithm [13] to minimize  $E(\mathbf{p})$ . The optimization process is to update the optimal increments  $\Delta\mathbf{p}$ , and then to update warping function  $\mathbf{W}(\mathbf{x};\mathbf{p})$ . By switching the role of template and target, the Hessian matrix could be pre-computed and re-used. Note that here the inverse incremental parameters is related to the regulation term.

$$E(\Delta\mathbf{p}) = \sum_x \| T(\mathbf{W}(\mathbf{x}; \Delta\mathbf{p})) - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \|^2 + \lambda R\left(\mathbf{p} + \frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} \Delta\mathbf{p}\right)^T \mathbf{K} R\left(\mathbf{p} + \frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} \Delta\mathbf{p}\right), \tag{8}$$

where  $\mathbf{p}'$  denotes the updated warping parameters, such as

$$\mathbf{W}(\mathbf{x}; \mathbf{p}') = \mathbf{W}(\mathbf{W}(\mathbf{x}; \Delta\mathbf{p})^{-1}; \mathbf{p}). \tag{9}$$

The function  $E(\Delta\mathbf{p})$  is nonlinear with respect to  $\Delta\mathbf{p}$ , which can be linearized by taking the first order Taylor expansion.

$$E(\Delta\mathbf{p}) \approx \sum_x \| T + \nabla T \frac{\partial\mathbf{W}}{\partial\mathbf{p}} \Delta\mathbf{p} - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \|^2 + \lambda \left[ R(\mathbf{p}) + \frac{\partial R}{\partial\mathbf{p}} \frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} \Delta\mathbf{p} \right]^T \mathbf{K} \left[ R(\mathbf{p}) + \frac{\partial R}{\partial\mathbf{p}} \frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} \Delta\mathbf{p} \right]. \tag{10}$$

According to definition of  $R(\mathbf{p})$ , we have  $\frac{\partial R}{\partial\mathbf{p}} = \mathbf{Q}$ . We set  $\mathbf{s}'$  as the reconstructed shape feature from  $\mathbf{p}'$  according to Eq (2). Combining Eq (9), it is computed as  $\mathbf{s}' = \mathbf{W}(\mathbf{s} - \mathbf{Q}\Delta\mathbf{p}; \mathbf{p})$ . Applying the chain rule, we have

$$\frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} = \frac{\partial\mathbf{p}'}{\partial\mathbf{s}'} \frac{\partial\mathbf{s}'}{\partial\Delta\mathbf{p}} = -\mathbf{Q}^{-1} \frac{\partial\mathbf{W}(\mathbf{x}; \mathbf{p})}{\partial\mathbf{x}} \mathbf{Q}. \tag{11}$$

Let  $\mathbf{B} = \frac{\partial R}{\partial\mathbf{p}} \frac{\partial\mathbf{p}'}{\partial\Delta\mathbf{p}} = -\frac{\partial\mathbf{W}(\mathbf{x}; \mathbf{p})}{\partial\mathbf{x}} \mathbf{Q}$ . The Hessian matrix is

$$\mathbf{H} = \sum_x \nabla T \frac{\partial\mathbf{W}}{\partial\mathbf{p}} \nabla T \frac{\partial\mathbf{W}}{\partial\mathbf{p}} + \lambda \mathbf{B}^T \mathbf{K} \mathbf{B}. \tag{12}$$

Finally, the solution of the above incremental parameter can be stated as:

$$\Delta\mathbf{p} = \mathbf{H}^{-1} \left( \sum_x (I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - T) - \lambda_c \mathbf{B}^T \mathbf{K} R(\mathbf{p}) \right). \tag{13}$$

### 3.3 Implementation

We design the proposed face alignment algorithm as an iteration process with local search and parameter optimization. We also notice that the PDM for shape features is trained on the normalized face examples. There is no rigid transformation factor in model parameter  $\mathbf{p}$ . To improve the model's adaptation to rotation and translation transformation, the local matching



points by K-PP are applied again. Given two sets of corresponding landmarks  $M = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{N_m}\}$  and  $D = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{N_m}\}$ , the rigid transformation parameters have been solved by the classic Iterative Closet Point algorithm. Thus we only give the conclusion here. The rotation matrix  $\mathbf{R}_\theta$  is calculated by a construction matrix  $\mathbf{C}$  and its singular value decomposition (SVD) [25].

$$\mathbf{C} = \frac{1}{N_m} \sum_{i=1}^{N_m} \mathbf{m}_i \mathbf{d}_i^T, \tag{14}$$

$$\mathbf{C} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}. \tag{15}$$

$$\mathbf{R}_\theta = \mathbf{V} \mathbf{U}^T. \tag{16}$$

And the translation matrix  $\mathbf{Tr}$  is computed by

$$\mathbf{Tr} = \frac{1}{N_m} \left( \sum_{i=1}^{N_m} \mathbf{m}_i - \sum_{i=1}^{N_m} \mathbf{R}_\theta \mathbf{d}_i \right). \tag{17}$$

The rigid transformation parameters are estimated at the beginning of each iteration process. The proposed algorithm is summarized as follows:

1. Feature Patch Generation: Given a template image, we obtain a set of feature patches around their landmarks.
2. K-PP Search: The proposed local K-PP search is performed for each feature patch.
3. Rigid Normalization: The rigid transformation parameters are computed by Eqs (16) and (17). The current landmarks are normalized by rigid transformation.
4. Local Constraint Construction: The weights  $w_i$  is calculated according to Eq (4). The regularization term about local matching constraint is constructed by Eq (6).
5. Parameter Optimization: The shape parameter  $\mathbf{p}$  is optimized by Eq (13). The features of current landmarks are updated.

Steps 2–5 are repeated until converged.

Note that after each iteration process, the local search regions are also changed with the current landmarks. In that case, the K-PP results and the regularization term are updated correspondingly.

## 4 Experimental Results

In the experiments, we select 263 face images from 80 subjects in the Cohn-Kanade Facial Expression Database [26] and 20 face images from 10 subjects captured by ourself. Their expressions include surprise, smile, sad, angry and neutral expression. Big facial deformation has more challenge, such as using a closed mouth to matching a big open mouth. So the faces we selected from the database are with the biggest extent of each expression of every person. All faces are annotated 80 landmarks manually. The training images are regularized by the positions of eye corners to  $256 \times 256$  pixels. To accommodate global lighting variation, the illumination intensity of face region is normalized in advance. In our experiments, the shape model remains 90% principle components. The regularization coefficient  $\lambda$  is set to 10.

We conduct two sets of experiments: the performance of local matching search and final alignment results. The target image has different expressions with the template image. Since the appearance changes, so the two image are mismatch. In the first set of experiments, our

goal is to find an optimal parameter  $K$  in the K-PP method to balance the similarity and the number of local matching. So the matching accuracy and matching number of different methods are exhibited to show the performance. In the second set of experiments, we compare our method with three different methods which could be used for non-rigid face alignment only using a single template. They are the Lucas-Kanade method (LK), the LK method combined with the nearest neighbor search (LK+NN), the LK method combined with the BBP (LK+BBP) and ours: LK with K-PP search (LK+K-PP).

#### 4.1 Performance of Local Matching Search

In order to evaluate the performance of local matching search, we compare the K-PP method with different parameters:  $k = 1$  (1-PP) which is also the BBP method,  $k = 2$  (2-PP),  $k = 4$  (4-PP),  $k = 6$  (6-PP), and the nearest neighbor search method (NN). 20 template and target image pairs are adopted in local matching experiments. Each image pair is the same person with different expressions. Training examples are not needed here. To test the robustness of each method, we randomly initialize the locations of landmarks by rotation and translation deformation and add noise in two ways: (a) rotated 5 degrees and randomly perturbed by Gaussian noise  $N(15, 10)$ , (b) rotated landmarks with 10 degrees and then perturbed landmarks by Gaussian noise  $N(10, 10)$ .

To quantitatively evaluate the accuracy of the algorithm, we calculate the Root Mean Square Error (RMSE) of testing results with manually annotated ground truth landmarks in Eq (18).

$$RMSE = \frac{\sum_{i=1}^N Dist(Result_i - Groundtruth_i)}{2 * N}, \tag{18}$$

where  $N$  is the number of landmarks and  $Dist(\cdot)$  is the Euclidean distance.

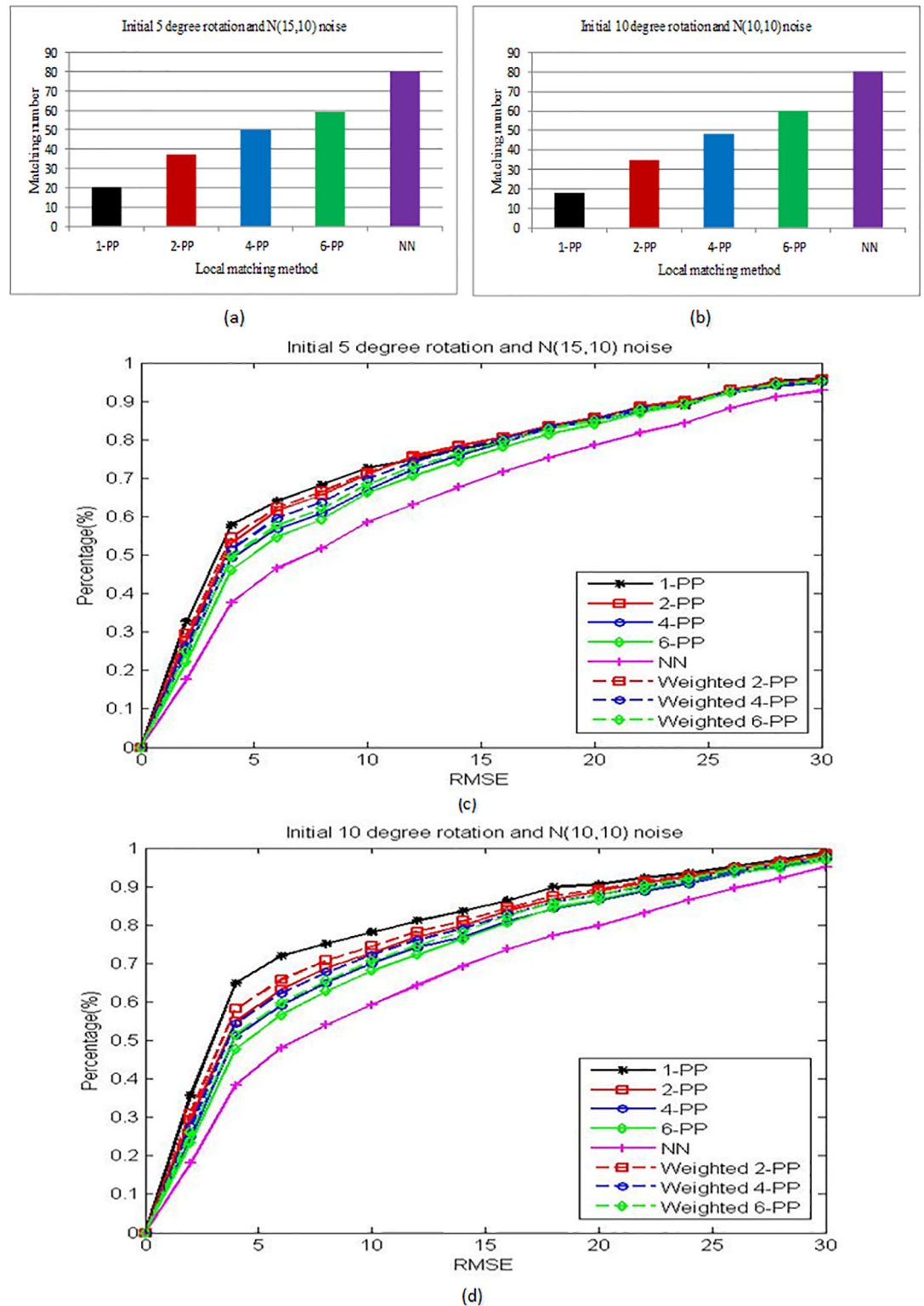
The number of local matching points is another criterion to evaluate the performance. As shown in Fig 2(a) and 2(b), we can see that the matching number increases when the parameter  $k$  becomes bigger. The 1-PP method finds the smallest number of matching points. The NN method can find the matching point for every landmark. So its matching number equals to the annotated landmark's number. The percentages of accumulative RMSE curves are drawn in Fig 2(c) and 2(d). We can see that the matching accuracy decrease when the  $K$  becomes bigger. The results by 1-PP method has the highest matching accuracy. And the results by NN method has the lowest matching accuracy. We also evaluate the effectiveness of weight matrix defined in Eq (4). The comparing results are shown as the solid curves and dashed curves respectively in Fig 2(c) and 2(d), which demonstrate that the weighted matching results have higher accuracy.

From the above experiment, the results verify that the matching number and the matching accuracy are two contradictory aspects in face alignment. But for the face alignment problem, both of them affect the final matching results. So in the following experiments, we use the weight matrix to balance them and choose  $k = 4$  as a trade-off parameter in the K-PP.

#### 4.2 Face Alignment Result

To demonstrate the generative of the proposed method, the appearance of testing images captured by ourself are Asian, which are totally different from the training database [26]. Fig 3 shows some examples of face alignment results. We compare our method with LK method and LK+NN method. The templates and target images with initial locations are shown in Fig 3(a) and 3(b). The template and target faces have very different expressions. So the facial appearances change a lot caused by geometric deformations. The results by different methods are





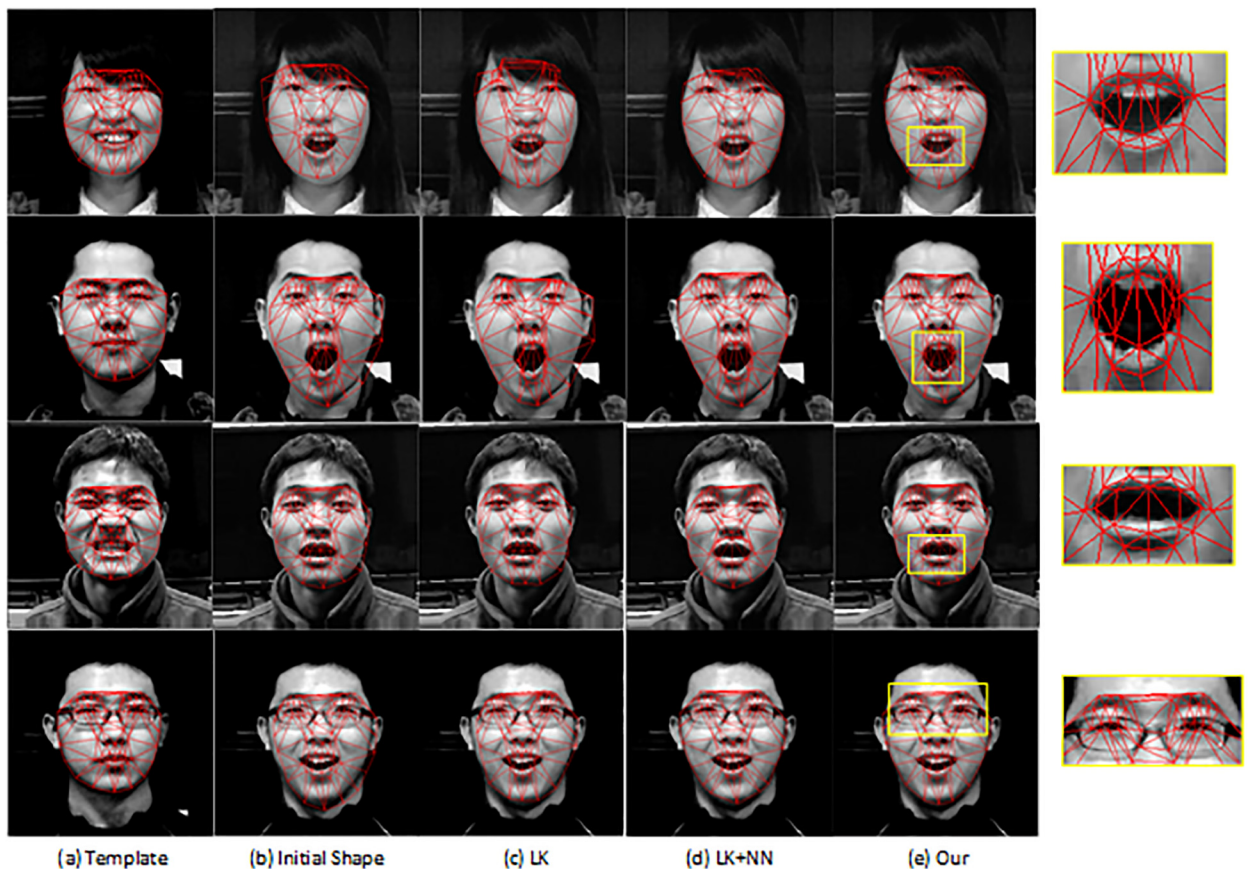
**Fig 2. Local matching search results of different methods including the nearest neighbor (NN) method, 1-PP (BBP), 2-PP, 4-PP and 6-PP method.** (a) and (b) show the average matching numbers when the initial location of landmarks are rotated 5 degrees and randomly perturbed by Gaussian noise  $N(15, 10)$ , and rotated 10 degrees and randomly perturbed by Gaussian noise  $N(10, 10)$ , respectively. (c) and (d) are the cumulative RMSE curves with corresponding initial locations of (a) and (b). The x-axis denotes the RMSE threshold and y-axis denotes the

percentage of test images which are less than each threshold. The solid curves are the K-PP results without weight or the weights are equal to 1. The dashed curves are the K-PP results with weights.

doi:10.1371/journal.pone.0159376.g002

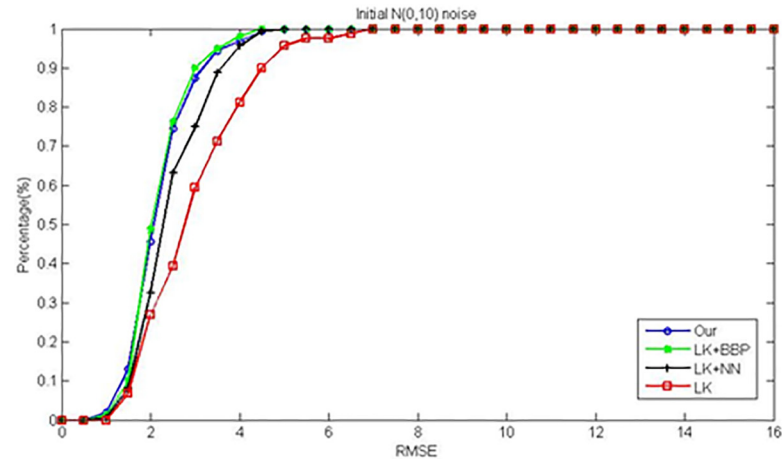
shown in Fig 3(c)–3(e). We can see that the LK method is based on the gradient descent algorithm, which is easy to be influenced by local texture leading to a local minimum solution. Better results can be obtained by combining with the NN search which consider both the holistic appearance and local features. But a few mismatch results by the NN search give negative influence to alignment results. For example, for the third person, some local patches of the mouth are mismatched to the nose. So the upper lip is dragged to the mouth by the LK+NN method. But our proposed method performs best which is robust to these outliers.

To quantitatively evaluate the accuracy of the algorithm, we calculate the Root Mean Square Error (RMSE) of testing results with manually annotated ground-truth. All 263 images in database [26] are tested. A leave-one-out cross-validation scheme is adopted throughout the quantitative experiments. When one image is used as target face, a corresponding image from the same person but with different expressions is selected as the target image. We compare four different methods: the proposed LK+4-PP method (Our), LK method, LK+BBP method and LK+NN method. Three kinds of noise are added to the positions of landmarks for initialization: (a) perturbed landmarks by Gaussian noise  $N(0, 10)$ , (b) rotated  $N(0, 5)$  degrees and

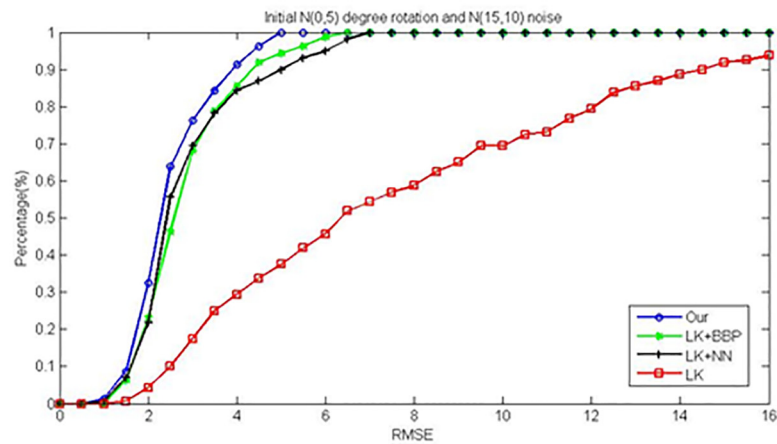


**Fig 3. Comparison of the face alignment using LK, LK+NN and our methods on real images.** (a) Template with annotated landmarks. (b) Target image with initial shape parameter. (c) The results of Lucas-Kanade method. (d) The results of LK method with the nearest neighbor search. (e) The results of our method, which is Lucas-Kanade method with KPP search ( $K = 4$ ).

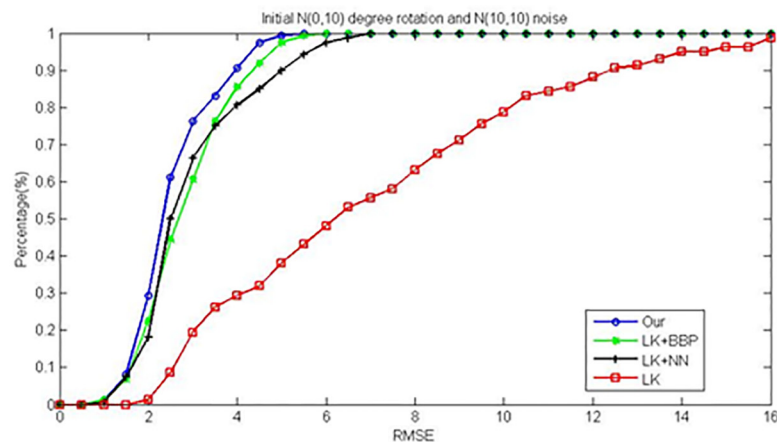
doi:10.1371/journal.pone.0159376.g003



(a)



(b)



(c)

**Fig 4. The cumulative RMSE curves with different initial parameters by four methods: Ours (LK +4-PP), LK+BBP, LK+NN and LK.** (a) The results with initial Gaussian noise  $N(0, 10)$ . (b) The results with initial rotated  $N(0, 5)$  degrees and randomly perturbed by Gaussian noise  $N(15, 10)$ . (c) The results with initial rotated  $N(0, 10)$  degrees and randomly perturbed by Gaussian noise  $N(10, 10)$ .

doi:10.1371/journal.pone.0159376.g004

**Table 1. The RMSE results for each facial component by different methods.**

	<i>N</i> (0, 5) rotation and <i>N</i> (15, 10) noise				<i>N</i> (0, 10) rotation and <i>N</i> (10, 10) noise			
	LK	LK+NN	LK+BBP	Our	LK	LK+NN	LK+BBP	Our
Eyes	5.68	3.65	3.46	<b>3.37</b>	5.30	3.77	3.53	<b>3.45</b>
Eyebrows	17.18	4.99	5.55	<b>4.79</b>	19.07	4.69	5.24	<b>4.58</b>
Nose	10.14	4.54	4.73	<b>4.16</b>	10.11	4.16	4.50	<b>3.88</b>
Mouth	16.83	8.29	6.69	<b>5.80</b>	19.51	8.37	6.83	<b>6.13</b>
Contour	17.36	7.40	7.36	<b>6.89</b>	19.46	6.78	7.56	<b>6.51</b>

doi:10.1371/journal.pone.0159376.t001

randomly perturbed by Gaussian noise  $N(15, 10)$ , (c) rotated  $N(0, 10)$  degrees and randomly perturbed by Gaussian noise  $N(10, 10)$ . Since the optimal parameter is the shape model parameter  $p$  in Eq (2), the initial landmarks are first projected to the low-dimensional linear space. So the actual initial position is normalized by the shape model.

The accumulative RMSE curves are shown in Fig 4. For the small initial noise in Fig 4(a), our method and LK+BBP method have better results. The LK+BBP method is a little better than ours. We analyse the reason is that, when the noise is small, the LK+BBP method is similar like the LK method with very good initial value. Then the optimization process can find the global optimal parameters. When given big noise, the comparing results are shown in Fig 4(b) and 4(c). Since there is no rigid transform parameter in the LK method, it plays very bad performance. But with the local search method in the LK+NN and our methods, the global rigid fitting could be applied naturally. Our method is not only robust to outliers but also have enough number of local matching prior, so our method is robust to the initial noise and has the best performance among the comparing methods.

Different facial components have different extents of deformations when doing expressions. We have noticed that the alignment results on facial local regions have much differences caused by various appearance changes. We record the average error in terms of the RMSE for eyes, eyebrows, nose, mouth and face contour in Table 1. The lower face region including mouth and contour may have bigger matching errors than upper face region. We can see our method still get the best results for each facial component. The average matching errors are lower than 7 pixels which outperforms those competing methods.

## 5 Conclusions

In this paper, a non-rigid face alignment method is proposed based on a single mismatched template. Considering two aspects about the matching rate and the matching number, a local matching method K-PP is presented based on mutual nearest neighbors searching. Furthermore, we employ the local matching features as a constraint for warping deformation parameters. The model fitting is demonstrated to be solved efficiently under the inverse compositional (IC) framework. Comparing to the conventional alignment method, the proposed method outperforms on the single template alignment.

Our contribution includes two main aspects: 1) A new local matching method K-PP is proposed which is robust to the complex facial deformation. 2) A formulation method for face alignment is proposed combining local and holistic matching. Both the flexible ability and robust ability are included which is useful for the non-rigid alignment problem when only given a single template. In the future work, more local features such as HOG features or Gaussian Filter will be used for illumination changing.

## Acknowledgments

This work was in part supported by the National Natural Science Foundation of China (61573274, 61379104), Postdoctoral Science Foundation of China (2015M582661), and Natural Science Basic Research Plan in Shaanxi Province of China (2015JQ6232).

## Author Contributions

**Conceived and designed the experiments:** YY SD.

**Performed the experiments:** YY ZC.

**Analyzed the data:** YY SD.

**Contributed reagents/materials/analysis tools:** YY.

**Wrote the paper:** YY.

## References

1. Mosquera L., and Castro J.L.: 'Performance of active appearance model-based pose-robust face recognition', *IET Computer Vision*, 2011, 5, (6), pp. 348–357 doi: [10.1049/iet-cvi.2010.0184](https://doi.org/10.1049/iet-cvi.2010.0184)
2. Sun Z., Lam K., et al.: 'Face Recognition with Multi-Resolution Spectral Feature Images', *Plos One*, 2013, 8, (2), pp. 1–12 doi: [10.1371/journal.pone.0055700](https://doi.org/10.1371/journal.pone.0055700)
3. Jones S., Dwyer D., Lewis M.: 'Learning Faces: Similar Comparator Faces Do Not Improve Performance', *Plos One*, 2015, 10, (1), pp. 1–21 doi: [10.1371/journal.pone.0116707](https://doi.org/10.1371/journal.pone.0116707)
4. Cootes T., Edwards G., and Taylor C.: 'Active appearance models', *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2001, 23, (6), pp. 681–685 doi: [10.1109/34.927467](https://doi.org/10.1109/34.927467)
5. Zhou M., Liang L., Sun J., and Wang Y.: 'AAM based face tracking with temporal matching and face segmentation', *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010, pp. 701–708
6. Liu X.: 'Video-based face model fitting using Adaptive Active Appearance Model', *Image and Vision Computing*, 2010, 28, (7), pp. 1162–1172 doi: [10.1016/j.imavis.2009.09.016](https://doi.org/10.1016/j.imavis.2009.09.016)
7. Walker K., Cootes T., and Taylor C.: 'Automatically building appearance models from image sequences using salient features', *Image and Vision Computing*, 2002, 20(5–6), pp. 435–440 doi: [10.1016/S0262-8856\(02\)00014-8](https://doi.org/10.1016/S0262-8856(02)00014-8)
8. Saragih J., and Goecke R.: 'Learning active appearance models from image sequences', *VisHCI*, 2006, 56, pp. 51–60
9. Gross R., Matthews I., and Baker S.: 'Generic vs. person specific active appearance models', *Image and Vision Computing*, 2005, 23(11), pp. 1080–1093 doi: [10.1016/j.imavis.2005.07.009](https://doi.org/10.1016/j.imavis.2005.07.009)
10. Pang Z., Wei C., et al.: 'Robust Eye Center Localization through Face Alignment and Invariant Iso-centric Patterns', *Plos One*, 2015, 10, (10), pp. 1–19 doi: [10.1371/journal.pone.0139098](https://doi.org/10.1371/journal.pone.0139098)
11. Ren S., Cao X., Wei Y. and Sun J.: 'Face alignment via regressing local binary features', *IEEE Trans. Image Processing*, 2016, 25, (3), pp. 1233–1245 doi: [10.1109/TIP.2016.2518867](https://doi.org/10.1109/TIP.2016.2518867)
12. Liu Q., Deng J. and Tao D.: 'Dual Sparse Constrained Cascade Regression for Robust Face Alignment', *IEEE Trans. Image Processing*, 2016, 25, (2), pp. 700–712 doi: [10.1109/TIP.2015.2502485](https://doi.org/10.1109/TIP.2015.2502485)
13. Baker S., and Matthews I.: 'Lucas-kanada 20 years on: a unifying framework', *Int'l J. Computer Vision*, 2004, 56(3), pp. 221–255 doi: [10.1023/B:VISI.0000011205.11775.fd](https://doi.org/10.1023/B:VISI.0000011205.11775.fd)
14. Yang Y., Liu Y., and Liu J.: 'Automatic face image annotation based on a single template with constrained warping deformation', *IET Computer Vision*, 2013, 7(1), pp. 20–28 doi: [10.1049/iet-cvi.2012.0094](https://doi.org/10.1049/iet-cvi.2012.0094)
15. Yang Y., Su Y., Cai D., and Xu M.: 'Nonlinear deformation learning for face alignment across expression and pose', *Neurocomputing*, 2016 (accepted).
16. Zhu J., Gool L. and Hoi S.: 'Unsupervised face alignment by robust nonrigid mapping', *Proc. International Conference on Computer Vision*, 2009, pp. 1265–1272
17. Cristinacce D., and Cootes T.: 'Feature detection and tracking with constrained local models', *Proc. British Machine Vision Conference*, 2006, pp. 929–938
18. Liang L., Xiao R., Wen F., and Sun J.: 'Face alignment via component-based discriminative search', *Proc. European Conference on Computer Vision*, 2008, pp. 72–85



19. Lowe D.: 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision*, 2004, vol. 60, no. 2, pp. 91–110 doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94)
20. Dalal N., and Triggs B.: 'Histograms of oriented gradients for human detection', *Int. Conf. on Computer Vision and Pattern Recognition*, 2005, 2, pp. 886–893
21. Pietikainen M., Zhao G., Hadid A., Ahonen T.: 'Computer vision using local binary pattern', Springer, 2011
22. Wang Y., Lucey S., and Cohn J.: 'Non-rigid object alignment with a mismatch template based on exhaustive local search', *Proc. IEEE Workshop on Non-rigid Registration and Tracking through Learning*, 2007, pp. 1–8
23. Detki T., Oron S., et. al.: 'Best-buddies similarity for robust template matching', *Proc. Computer Vision and Pattern Recognition*, 2015, pp. 2021–2029
24. Cootes T., Taylor C., and Graham J.: 'Active shape models-their training and application', *Computer Vision and Image Understanding*, 1995, 61(1), pp. 38–59 doi: [10.1006/cviu.1995.1004](https://doi.org/10.1006/cviu.1995.1004)
25. Golub, Gene H., Van C., Charles F.: *Matrix Computations* (3rd ed.), Johns Hopkins, 1996
26. Kanade T., Cohn J.F., and Tian Y.: 'Comprehensive database for facial expression analysis', *Automatic Face and Gesture Recognition*, 2000, pp. 46–53