IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 27, 2020, accepted March 20, 2020, date of publication March 27, 2020, date of current version April 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983722

Phased Groupwise Face Alignment

GANG ZHANG[1], HAN QIN[1], YUDING KE[1], JIANSHENG CHEN[2], AND YANMIN GONG[3]
[1]College of Software, Shenyang University of Technology, Shenyang 110023, China
[2]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[3]Electrical and Computer Engineering, The University of Texas at San Antonio, San Antonio, TX 78249, USA

Corresponding author: Gang Zhang (zhang_gang_1973@yahoo.com)

This work was supported by the National Natural Science Foundation of China under Grant 61372176.

ABSTRACT A face does not only have rigid variations but also non-rigid distortions, which has influenced the performance of groupwise face alignment. A novel method for groupwise face alignment which considers both rigid variations of a face and non-rigid distortions was presented in the paper. The process for groupwise face alignment was divided into two stages, *i.e.* affine transformations and non-rigid distortions. At the stage of the affine transformations, the key points of a face were categorized into five groups and the affine transformations were used for each group of the key points. At the stage of the non-rigid distortions, a novel method was used for all of the key points in a face. Two stages were independent of each other, and moreover, iterations were made in each stage. Besides, the results from the stage of the affine transformations were used as the input of the non-rigid distortion stage. Experiments show that the method for groupwise face alignment in the paper is better than that only considering global affine variations, and is also better than that considering global affine variations and local non-rigid distortions. The method in the paper can be used as a novel method for groupwise face alignment.

INDEX TERMS Face analysis, face alignment, joint alignment, groupwise alignment.

I. INTRODUCTION

Joint face alignment which is also known as groupwise face alignment considers multiple face images of the same person or different persons. It can improve alignment performance of these faces by using the constraints and compensation among them. Certainly, great progresses in face alignment had been made due to use of prior information and auxiliary models, especially after the emergence of deep learning [1]–[4]. However, only a face is usually available during the alignment, and the lack of information restricts the results of face alignment. Comparatively speaking, groupwise face alignment provides a feasible way to solve this problem. After all, there are more faces available during the alignment. The method using groupwise face alignment can provide strong support for the applications, for examples, face analysis, face synthesis.

Researches on groupwise face alignment mainly focus on the contents, *i.e.* mappings among faces, face representation, and groupwise optimization.

The previous methods for groupwise alignment consider affine variations among the same objects in the images [5], and is only suitable for groupwise alignment among the same objects obtained by making the in-plane affine transformations. As a face does not only have global affine variations but also local non-rigid distortions, these methods are not suitable for faces. Cootes *et al.* [6] presented a classical method for groupwise face alignment considering both global affine variations and local non-rigid distortions. Generalized procrustes analysis was used to superimpose the faces in images and then the mean face was obtained. Piecewise affine distortion was used to map the texture of each face to the mean face. Then the residual distribution was obtained by using the textures of all faces. The key points of all faces were optimized by using the residual distribution. During groupwise optimization, the downhill simplex of Nelder-Mead was used to estimate global affine transformations, and local transformations and gradient-based search were used to determine the results of local non-rigid distortions.

Face representation is one of the most important factors in groupwise face alignment, and it focuses on the representation and organization of faces. A face mainly contains shape and texture information. During groupwise face alignment, the shape information is mainly considered but the texture information is usually used to construct the constraints for the optimization [7]. The methods for face representation can be categorized into two kinds. The first kind of methods use the

The associate editor coordinating the review of this manuscript and approving it for publication was Madhu S. Nair.

positions of the key points in a face to represent its shape and use a matrix to store the shapes of all faces [8]. Each row or each column corresponds to the shape of a face in the matrix. The second kind of methods use the parameters of facial pose, *i.e.* yaw, pitch and roll, and those of local distortions to represent the shape of a face [9]. Comparatively speaking, as the first kind of methods considers the positions of the key points, they need a large storage space. But as the second kind of methods takes into account the set of parameters, they only need a small storage space. The first kind of methods for face representation is called non-parametric representation here, and the second kind of methods for face representation is called parametric representation. Here the non-parametric representation is used.

Optimization is also one of the most important factors in groupwise face alignment for non-parametric representation or parametric representation. During the optimization, both global affine variation and local non-rigid distortion of each face need to be considered in multiple faces [10]–[12]. For the non-parametric representation, the reference face is used to constrain the optimization, and moreover, it is updated dynamically during the optimization. For the parametric representation, the optimization does not act on the key points but on the transformation parameters which contain the parameters for both the global affine transformations and the local non-rigid distortions. So, the problem about the optimization is transformed into that about estimating the set of parameters. For the nonparametric representation, the Newton descent method and its variants are usually used to seek for the least square solution which can be obtained iteratively. For the parametric representation, deep learning and regression are usually used.

The termination condition for optimization is one factor which needs to be considered in the optimization. In the initial iterations, the key points or the set of parameters will be updated greatly. Afterwards, they will be updated slowly. When the intermediate results are close to or up to the actual values, the shrink case will come if the iteration continues to be made. All of the key points in a face will shrink to the centroid of the face. As the termination condition for optimization, the maximum number of iterations is usually used or that amount of variations between the positions of the key points or the parameter values are less than the specified thresholds after two consecutive iterations. For the non-parametric representation, the minimum rank of matrix can be used as the termination condition. However, it needs to be noted that it is usually difficult to obtain good results to use the rank of 1 as the termination condition.

It is not a vital factor to select a reference face during groupwise alignment, as long as the positions of the key points in the reference face does not deviate from the actual positions of the key points in multiple faces too much [6]. During groupwise alignment, the reference face is used to constrain the optimization of the key points in multiple faces at each iteration and then it will be updated dynamically. When the positions of the key points in the initial reference face deviate from the actual positions of the key points in multiple faces too much, the performance of groupwise face alignment will be affected greatly [13], [14].

TABLE 1. Comparisons of methods for groupwise face alignment.

Method	Considered factors
Learned-Miller, 2006[5]	In-plane affine variations
Cootes *et al.*, 2010[6]	Global affine variations and local non-rigid distortions (residual distribution)
Zhao et al., 2011[7]	Global affine variations and local non-rigid distortions (subspace)
Anderson et al., 2014[13]	A reference face
Liao et al., 2014[14]	A reference face
Peng et al., 2018[8]	Non-parametric representation and optimization

Here, comparisons of some methods for groupwise face alignment are shown in Table 1. The novel method for groupwise face alignment in the paper is different from that only considering the global affine variations, and it considers the non-rigid distortion of each face during the optimization. Besides, it uses a novel method for non-rigid distortions. It is also different from the method for groupwise face alignment considering global affine variations and local non-rigid distortions. It divides the key points of each face into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth. The affine transformations are made for each group, respectively. Then the non-rigid distortions are made for all key points in each face. *As far as we know, it is the first time to group the key points of faces and then introduce the grouped results into groupwise face alignment.*

The contributions of this paper are as follows. First, both the affine variations and non-rigid distortions are considered, and moreover, regarded as two independent stages during groupwise face alignment. Besides, a novel method for non-rigid distortions is used. Second, the case that the key points of a face are regarded as a whole, and moreover, global affine transformations and non-rigid distortions are made iteratively, is analyzed. Besides, it was found that if the constraints among the groups of the key points are too strict, the performance of the affine transformations will decrease. So, the key points of each face are grouped, and the affine transformations are made for each group, respectively. Afterwards, non-rigid distortions are made for all the key points of each face. Third, the reference face is transformed into that related to multiple faces. Then it is updated iteratively at the stage of the affine transformations, and it is used as a constraint in the stage. Fourth, for the stage of the affine transformations and non-rigid distortions, the maximum number of iterations is used as the termination condition for the iterative optimization to solve the shrinking problem about the alignment result of each face.

Section II describes the principle of groupwise face alignment and the symbols are normalized in this paper. Section III describes the method for face alignment. Section IV describes the method for groupwise face alignment which contains face representation, optimization, and termination conditions. Section V is experiments and performance analysis. Section VI concludes the paper.

II. PRINCIPLE OF GROUPWISE FACE ALIGNMENT AND SYMBOLIC NORMALIZATION

A face is different from rigid objects in that it does not only contain rigid variations but also non-rigid distortions. Besides, it was found that if the constraints among different parts of a face are too strict, the performance of the affine transformations will decrease. So, a novel method for groupwise face alignment is presented in the paper. It still optimizes the alignment result of multiple faces, but it regards the affine variations and the non-rigid distortions as two independent stages. At each stage, the optimization is performed iteratively, respectively. Besides, the reference face is updated dynamically during the optimization. At the stage of the affine transformations, the key points of each face are divided into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth, and the affine transformations are made for each group, respectively. At the stage of the non-rigid distortions, all the key points of a face are considered. At each stage, the maximum number of iterations is used as the termination condition of the optimization. The description of groupwise face alignment is shown in Figure 1.

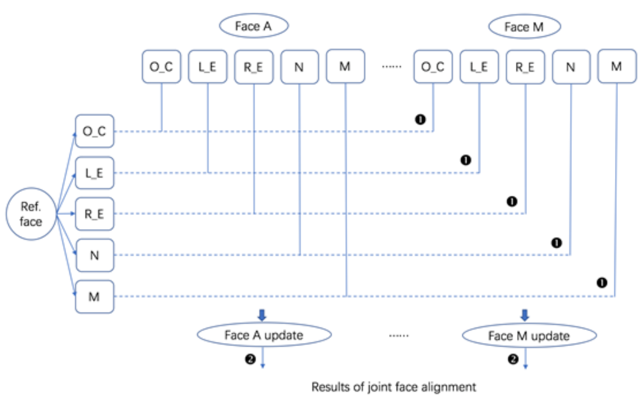


FIGURE 1. Groupwise face alignment O_C: outer contour, L_E: left eye, R_E: right eye, N: nose, M: mouth 1. Affine transformations for each group 2. Non-rigid distortions for all the key points.

Here the symbols in the paper will be normalized. Suppose that the M faces are represented as $\mathrm{F} = \{f_i(x, y) \mid i \in [1, \mathrm{M}]\}$, the positions of the C key points are represented as $\mathrm{SF} = \{f_i^t(x, y) \mid i \in [1, \mathrm{M}] \wedge t \in [1, \mathrm{C}]\}$ in the M images, the alignment results of the M images are represented as $\mathrm{GF} = \{g_j\ (f_i(x, y)) \mid i \in [1, \mathrm{M}] \wedge j \in [1, \mathrm{N}]\}$ after the j-th affine transformations, the alignment results of the M images are represented as $\mathrm{LF} = \{l_j\ (f_i(x, y)) \mid i \in [1, M] \wedge j \in [1, \mathrm{N}]\}$ after the j-th non-rigid distortions, the reference face used at the j-th iteration is represented as ref_j and $j \in [1, \mathrm{N}]$ where N is the maximum number of iterations, the grouping functions of the C key points are represented as $\mathrm{GR}=\{\mathrm{G}_k\ (\mathrm{T}) \mid \mathrm{G}_k=\{\mathrm{OuterContour, LeftEye, RightEye, Nose, Mouth}\}\}$, the key points of a face is represented as T where T is either a face in multiple faces or a reference face.

III. FACE ALIGNMENT

Groupwise face alignment is usually used as the method for optimizing the results of face alignment. Face alignment is first made for each face in multiple faces. Then the constraints among multiple faces are used to optimize the alignment results further. For face alignment in the wild, local methods for face alignment have received much attention. Classic methods contain constrained local model (CLM), supervised descent method (SDM), convolution neural network (CNN), and *etc*. [15]–[18]. Here the SDM is used. Certainly, the methods using deep learning can also be used. No matter which method is used, the alignment result of a face can be obtained.

For the i-th face in multiple faces, suppose that the C key points of a face are represented as $\mathrm{d}_i(u)$, $\mathrm{d}_i(u) = \{(x_i, y_i) \mid i \in [1, C]\}$ and u is the positions of the C key points, the nonlinear function for feature extraction is represented as h and the scale-invariant feature transform descriptor is used as the function, the feature vector of 128∗C dimension by using the function h for the positions of ground truths of the C key points in the i-th face, $\emptyset_*$, can be obtained. Then the minimum Δu can be estimated by using the equation (1).

$$\Delta u = \min_{\Delta u} \| h(d(u_+\Delta_u)) - \emptyset_* \| \tag{1}$$

The u_0 is only known during the alignment and the Δu and $\emptyset_*$ are unknown. Besides, it is not practical to solve Δu at one time. So, the Δu can be solved iteratively by using regression and can be represented as

$$\mathrm{u}_k = u_{k-1} - 2\mathrm{H}^{-1}\mathrm{J}_h^{\mathrm{T}}(\emptyset_{k-1} - \emptyset_*) \tag{2}$$

Here the u_{k-1} and u_k are the positions of the C key points in the i-th face after two consecutive iterations, respectively, the H and J_h are the Hessen matrix and the Jacobian matrix, respectively, the $\emptyset_{k-1}$ is the feature vector of 128∗C dimension which is extracted from the C key points after the k-1-th iteration. Let $\mathrm{R}_{k-1} = -2\mathrm{H}^{-1}\mathrm{J}_h^{\mathrm{T}}\emptyset_{k-1}$ and $\mathrm{b}_{k-1} = 2\mathrm{H}^{-1}\mathrm{J}_h^{\mathrm{T}}\emptyset_*$, then the equation (2) can be represented as followed.

$$\mathrm{u}_k = u_{k-1} + \mathrm{R}_{k-1}\emptyset_{k-1} + \mathrm{b}_{k-1} \tag{3}$$

Thus, the problem about face alignment is transformed into that about solving a set of parameters R_k and b_k by using regression.

IV. GROUPWISE FACE ALIGNMENT

A. FACE REPRESENTATION OF GROUPWISE ALIGNMENT

Here the non-parametric representation is used in groupwise face alignment. The shapes of faces are considered and stored by using a matrix. The shape of each face is described by using the positions of the key points which are organized in an order. First, the x component is stored and then the y component is done. Each row or each column corresponds to the shape of a face in the matrix.

Suppose that only a single face is contained in each image, the i-th face is represented as $f_i\ (x, y)$, the positions of the k-th key point are represented as $(x_{i,k}, y_{i,k})$ in the i-th image and

$k \in [1, C]$, the number of the key points is C, then the shapes of multiple faces can be represented as followed.

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,C} & y_{1,1} & \cdots & y_{1,C} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{M,1} & \cdots & x_{M,C} & y_{M,1} & \cdots & y_{M,C} \end{bmatrix} \tag{4}$$

It is advantageous to use a matrix to store the shapes of multiple faces. For examples, the methods for optimization are easy to be selected as well as the termination conditions. The shapes in the equation (4) can be updated dynamically during groupwise alignment, and moreover, they can be represented as (5), as shown at the bottom of the next page. In addition, the rank minimum of the matrix can be used as the termination condition. It needs to be noted that for multiple faces from different persons or the same person, good results will not be obtained by using the termination condition that the rank is equal to 1 in the wild [19].

B. OPTIMIZATION USING GROUPWISE ALIGNMENT

In view of the fact that a face has rigid variations and non-rigid distortions, groupwise face alignment regards rigid variations and non-rigid distortions as two independent stages in the paper. At the stage of the affine transformations, the key points of a face are divided into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth. The affine transformations are made iteratively for each group, and the reference face is updated dynamically. At the stage of the non-rigid distortions, all of the key points in a face are used as a whole, and but local differences between the key points are considered in different faces. The complete algorithm using groupwise optimization is detailed as follows.

The process for optimization is shown in Figure 2.

First, the key points of the reference face are divided into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth, as shown in Figure 3. The five groups of the reference face are represented as

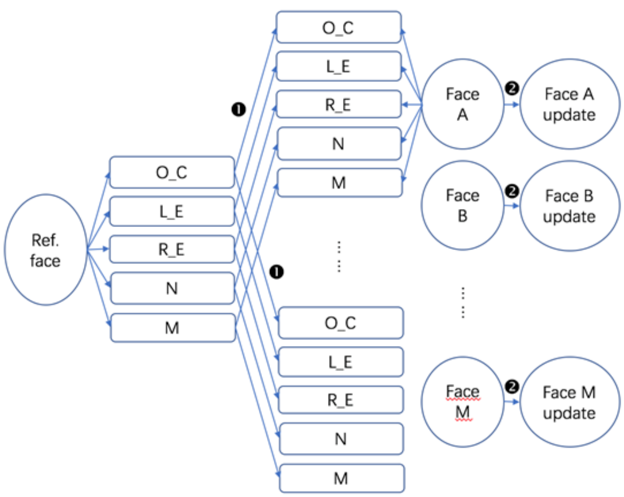


FIGURE 2. **Process of optimization O_C: outer contour, L_E: left eye, R_E: right eye, N: nose, M: mouth 1. Affine transformations for each group 2. Non-rigid distortions for all key points.**

Algorithm 1 Groupwise Alignment

```
Input: the sets of key points by using the method in III for multiple faces
Obtain the reference face from the training set of faces
 for the set of key points in each of multiple faces
     Divide them into five groups
     for each group
        Let sp the set of key points in the group
        Let ref the corresponding set in the reference face
        Find affine transformations from ref to sp
        Transform ref by using the affine transformations
     end for
 end for
Superimpose the transformed sets of key points in all faces for each group
Recomputed the mean for each group and then construct a new reference face
Compute 2 norm between the mean face and the previous reference face
   if value is larger than a threshold &&
      the number of iteration is less than the maximum number of iterations
      Repeat the above process
   else
      terminate the iterations
   end if
Perform non-rigid distortion for the transformed sets of all key points
Output: the sets of key points by using groupwise optimization
```

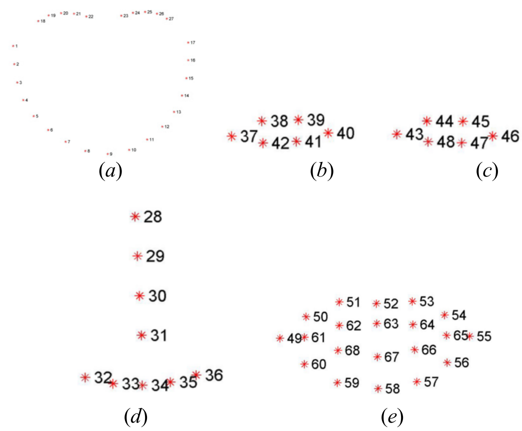


FIGURE 3. **Five groups of the key points.**

OuterContour(ref_j), LeftEye(ref_j), RightEye(ref_j), Nose(ref_j), and Mouth(ref_j), respectively, when the j-th iteration is made. The five groups of the i-th face in multiple faces are represented as OuterContour(SF_i), LeftEye(SF_i), RightEye(SF_i), Nose(SF_i), and Mouth(SF_i), respectively. For OuterContour(ref_j) and OuterContour(SF_i), the affine transformations tr from OuterContour(ref_j) to OuterContour(SF_i) are determined by the equation (6) which uses Random Sample Consensus. After the processes to estimate the parameters for the affine transformations are made from OuterContour(ref_j) to its corresponding groups in SF, the set of the parameters for the affine transformations can be determined. By means of the set of parameters, the OuterContour(ref_j) is transformed to obtain the sets of points from outer contours of the faces in GF after the j-th affine transformation. Grouping and mapping are made for multiple faces. Then the generalized procrustes analysis is used for the sets of points, and the mean point set can be obtained and be used as the reference, as shown in the equation (7).

$$tr : \mathrm{OuterContour}(ref_j) \rightarrow \mathrm{OuterContour}(\mathrm{SF}_i) \quad (6)$$

$$ref_{j+1} = mean(superimpose(OuterContour(SF))) \quad (7)$$

Here $mean()$ is the mean operator, $superimpose()$ is the operator for superimposing and is implemented by using the generalized procrustes analysis, ref_j and ref_{j+1} are the mean which are used in the j-th and j+1-th affine transformations, respectively. These steps are iterated until the maximum number of iterations, and the affine transformations of the group are completed. The above process is made for LeftEye(ref_j) and LeftEye(SF_i), RightEye(ref_j) and RightEye(SF_i), Nose(ref_j) and Nose(SF_i), and Mouth(ref_j) and Mouth(SF_i), respectively. The affine transformations of all groups are completed.

After the stage of the affine transformations, there are five groups for each face in multiple faces, *i.e.* outer contour of a face, left eye, right eye, nose and mouth. These groups are linked according to the order of points of the faces in SF. Besides, the x components are stored first and then the y components are done for the shape of each face. Thus, each row or each column in the matrix corresponds to the shape of a face.

Second, non-rigid distortions are made for all key points of faces which are the results GF of the affine transformations. It considers the sets of points GF and SF in multiple faces. For any face i, it seeks for the mapping between GF_i and SF_i. It is meaningless to make complete fitting between GF_i and SF_i. After all, SF_i itself can have local deviations. In contrast, it is more valuable to use the k pairs of points. One reason is that these points in GF_i are close to the corresponding points in SF_i. The other reason is that these points in GF_i are constrained by facial shape models. Thus, the problem of non-rigid distortions is transformed into that about seeking for the k pairs of the closest points between GF_i and SF_i, and looking for the mapping between the sets of the closest point pairs. Afterwards, the parameters for the mapping will be used for GF_i to compute the results of non-rigid distortions.

For the sets of points, GF_i and SF_i, the points which are the closest to the corresponding points in SF_i are found in GF_i. These points from GF_i and SF_i are used to construct the set of the closest point pairs CSP_i, as shown in the equation (8). The parameters for mapping between GF_i and SF_i are found by using CSP_i. Then the GF_i is transformed further by using these parameters to obtain the results of non-rigid distortions LF_i, as shown in the equation (9).

$$CSP_i = \mathrm{knearest}(GF_i, SF_i) \quad (8)$$

$$LF_i = nrtr(GF_i, CSP_i) \quad (9)$$

The above process is repeated for each face in GF and SF to obtain the set of the distortion results LF.

It should be noted that affine transformations and non-rigid distortions are regarded as two independent stages, and iterations are made at each stage, respectively. At the stage of the affine transformation, the constraints among multiple faces are imposed by using the reference face. At the stage of the non-rigid distortions, the personalized characteristics of faces are considered and the optimization is constrained by facial shape models.

C. TERMINATION CONDITIONS OF GROUPWISE ALIGNMENT

It was found that for the groupwise face alignment only considering global affine variations, if the alignment results of a face are close to or even up to their actual positions during the optimization, the shrinking case will come when the iterations continue to be made. All of the key points will shrink to the centroid of the face. It can solve the problem to group the key points of a face. Besides, it can also solve the problem to set the termination condition for constraining the optimization. Here, besides face grouping, the termination condition will be used to constrain the iterative optimization of two stages, *i.e.* the stages of the affine transformations and the non-rigid distortions. One way to set the termination condition is to impose the constraints of the rank minimization of the matrix according to the characteristics of the face representation in section IV. The other way is to use the maximum number of iterations.

D. DISCUSSION

Compared with alignment for a single face, groupwise face alignment has many advantages. The previous methods for groupwise optimization only impose groupwise constraints on common information of faces, and here extension of groupwise constraints is made with individual information of each face considered. Besides, the defects that global variations and local distortions are used in order for all the

$$\begin{bmatrix} x_{1,1}+\Delta x_{1,1} & \cdots & x_{1,C}+\Delta x_{1,C} & y_{1,1}+\Delta y_{1,1} & \cdots & y_{1,C}+\Delta y_{1,C} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{M,1}+\Delta x_{M,1} & \cdots & x_{M,C}+\Delta x_{M,C} & y_{M,1}+\Delta y_{M,1} & \cdots & y_{M,C}+\Delta y_{M,C} \end{bmatrix} \quad (5)$$

key points are considered. Thus, when global transformations are performed, the key points of a face are grouped and affine transforms are made in each group. In this way, the case that all the key points are constrained too strictly can decrease. When local non-rigid distortions are made, all the key points are considered. This way, it can be guaranteed that local distortions are considered and global constraints can be imposed.

V. EXPERIMENTS AND PERFORMANCE ANALYSIS

The experiments are carried out from the aspects as follows. First, the method of face alignment is used in the paper. The groupwise face alignment only considering global affine variations is used for the optimization of the alignment results. All of the key points are used when the affine transformations are made. Second, the method of face alignment is used in the paper. The groupwise face alignment considering global affine variations and local non-rigid distortions is used for the optimization of the alignment results. The global affine transformations and local non-rigid distortions are made in order at each iteration, and all of the key points are used. Third, the method of face alignment is used in the paper. The groupwise face alignment in section IV is used to optimize the results of the face alignment. The key points are grouped and the affine transformations are made for each group. The results of the affine transformations will be used in the non-rigid distortions and all of the key points are considered. Fourth, the performance of three methods is compared to determine the optimal method for groupwise face alignment. The reason that the comparisons are made between the previous two methods and the method in the paper is that the first method is mainly used in the previous methods for groupwise alignment and the second method is mainly used in the current methods for groupwise alignment.

The faces for testing come from the Helen face database [20]. The front faces and the faces which are close to front are selected from the database to ensure that all key point of each face are visible. Thus, the effect of facial pose variations on three methods can be eliminated. In addition, the faces used in groupwise face alignment need to meet the requirements as follows. First, local deviations of the results of face alignment only occur in a small number of faces. Second, for each of the key points, most of the faces are aligned successfully or the alignment results of most faces are close to their actual positions. By using these faces, the optimal method for groupwise face alignment will be determined.

The subset of faces for testing are selected from the Helen face database. Each subset contains the six faces and each face is selected randomly. For each subset, the AdaBoost classifiers are used to determine the external rectangle of each face, and then the method of face alignment in the paper are used to perform alignment of the face. For the groupwise face alignment only considering global affine transformations, the groupwise face alignment is made iteratively by using the equations (6) and (7). All of the key points are considered. For the groupwise face alignment considering global affine variations and local non-rigid distortions, all of the key points are considered, and the affine transformations and non-rigid distortions are made in order at each iteration. At the stage of the affine transformations, the optimization is made by using the equations (6) and (7), and the non-rigid distortions are made by using the equations (8) and (9). For the groupwise face alignment in the paper, the key points of a face are divided into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth, and the equations (6) and (7) are used iteratively for each group. Then the results of the affine transformations are used for the stage of non-rigid distortions. At the stage, all of the key points are used, and the equations (8) and (9) are used iteratively. It should be noted that the affine transformations and non-rigid distortions are independent of each other. Besides, the maximum number of iterations is used as the termination condition.

The visual effect of three methods for groupwise face alignment is shown in Figure 4, groupwise face alignment only considering global affine variations, that considering global affine variations and non-rigid distortions, that of IV. Afterwards, statistical analyses are made on the set of faces for testing by using the equation (10) and (11) where the square sum of the difference between the key points of faces is used as a measure, err_i is the measure of the i-th face in the results of groupwise face alignment, $rst_{i,j}$ is the result of the j-th key point in the i-th face, $rst_{gt,j}$ is the ground truth of $rst_{i,j}$. The statistical value uses the average result of groupwise face alignment for the front or close front faces in the Helen face database.

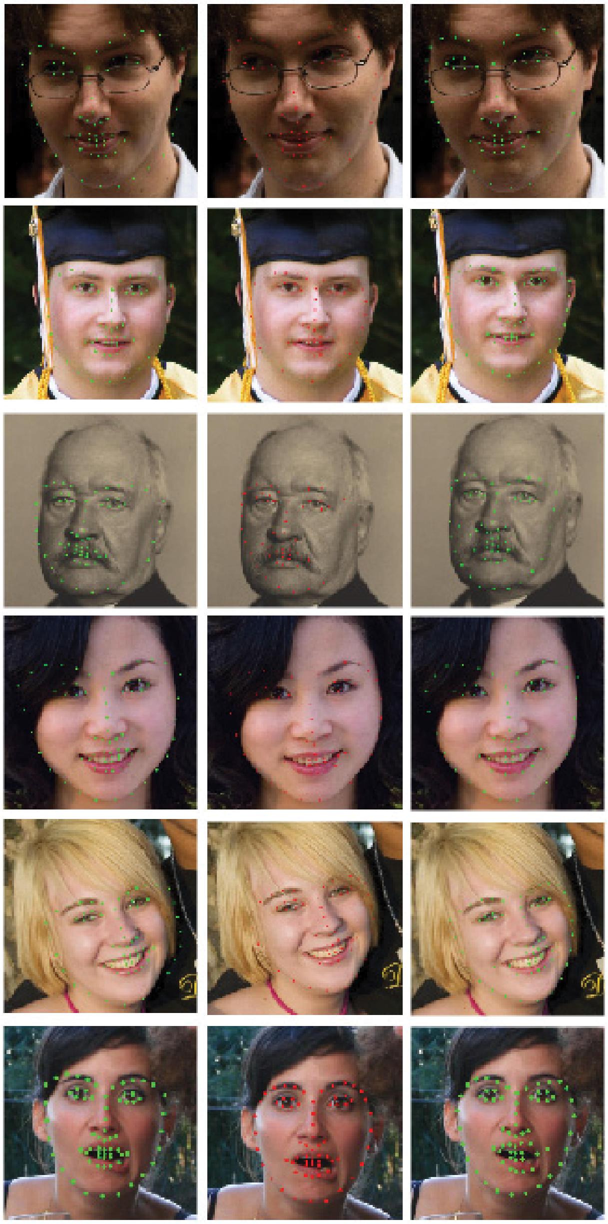
$$\mathrm{err}_i = \sum_{j=1}^{c} (rst_{i,j} - rst_{gt,j})^2 \tag{10}$$

$$\mathrm{err} = \frac{1}{N} \sum_{i=1}^{N} err_i \tag{11}$$

The statistical values of three methods for groupwise face alignment are shown in Table 2. It can be found that the method for groupwise face alignment considering grouping of the key points obtains the optimal results. The effect of three methods on local key points of faces is shown in Figure 5. It can be found that when the method for groupwise face alignment only considering global affine variations is used, the key points which are close to the corner of right eye sometimes deviate. When the method for groupwise face alignment considering global affine variations and non-rigid distortions is used, better results can be obtained. When the method for groupwise face alignment in section IV is used, the best results can be obtained. This means that it is valuable

TABLE 2. Statistics of alignment performance of three methods.

Method	Only considering global affine variations	Considering global affine variations and non-rigid distortions	Considering grouping and non-rigid distortions
Measure	90507.549	90199.883	88728.075



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

FIGURE 4. **Visual effect of three methods for groupwise face alignment (a)Only considering global affine variations (b)Considering global affine variations and non-rigid distortions (c)Considering grouping and non-rigid distortions.**

to consider global affine variations and non-rigid distortions. Besides, it is advantageous to group the key points of a face and perform the affine transformations for each group.

The method for groupwise face alignment in the paper considers the grouping of the key points and non-rigid distortion, and regards affine variations and non-rigid distortions as independent stages. At each stage, the optimal solution can be obtained iteratively, and the maximum number of iterations is used as the termination condition of the optimization. Here the effect of the number of iterations on groupwise face alignment will be analyzed. It was found that when the maximum number of iterations is set to 5 in each of

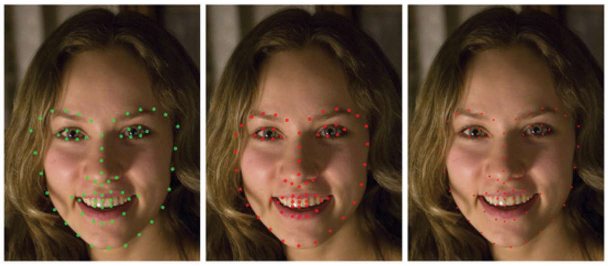


FIGURE 5. **Local results of using three methods for groupwise face alignment.**

two stages, the performance of groupwise face alignment measured by using the equations (10) and (11) has tended to be stable. When the maximum number of iteration varies from 5 to 25 with 5 as an unit, the performance of groupwise face alignment has slightly varied. So, the maximum number of iterations is set to 5 in the experiments.

From the perspective of time complexity, the first method is the lowest, and the method in the paper is slightly higher than the second method. When the method for groupwise face alignment in IV is used, some cases of failure will occur. It was found that when the non-rigid distortion cannot describe the non-linear variations of faces, the problem occurs. In addition, failures mostly occur in the points of eyebrows, which is caused by the case that the points from eyebrows and outer contour of a face are considered together in Figure 3(*a*).

VI. CONCLUSION

A novel method for groupwise face alignment is presented in the paper. It considers affine variations and non-rigid distortions of faces, and regards groupwise face alignment as two stages, *i.e.* affine variations and non-rigid distortions. At the stage of the affine variations, the key points of a face are divided into five groups, *i.e.* outer contour of a face, left eye, right eye, nose and mouth, and the affine transformations are made for each group. Afterwards, the non-rigid distortions are used for all of the key points in a face. Two stages are independent to each other, and the iterations are made at each stage. At the stage of the affine variations, the grouping of the key points is made and the affine transformations are made for each group. At the stage of the non-rigid distortions, the non-rigid variations of faces are considered. The method for groupwise face alignment in the paper is better than that only considering global affine variations, and is also better than that considering global affine variations and non-rigid distortions. In addition, it benefits from the grouping of the key points and avoids the shrinking case of the alignment results after groupwise face alignment. In the latter works, the effective method for fitting non-rigid distortions of faces will be studied further.

ACKNOWLEDGMENT

The authors would like to thank good suggestions from anonymous reviewers. They would also thankful to other researches for providing open source algorithms and the image sets.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[2] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (And a dataset of 230,000 3D facial landmarks)," in *Proc. ICCV*, Venice, Italy, Jun. 2017, pp. 1021–1030.

[3] H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen, "Heterogeneous face attribute estimation: A deep multi-task learning approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2597–2609, Nov. 2018.

[4] F. J. Chang, A. T. Tran, T. Hassner, I. Masi, R. Nevatia, and G. Medioni, "Deep, landmark-free FAME: Face alignment, modeling, and expression estimation," *Int. J. Comput. Vis.*, vol. 127, nos. 6–7, pp. 930–956, Jun. 2019.

[5] E. G. Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 236–250, Feb. 2006.

[6] T. F. Cootes, C. J. Twining, V. S. Petrovic, K. O. Babalola, and C. J. Taylor, "Computing accurate correspondences across groups of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1994–2005, Nov. 2010.

[7] C. Zhao, W. K. Cham, and X. Wang, "Joint face alignment with a generic deformable face model," in *Proc. CVPR*, Washington, DC, USA, Jun. 2011.

[8] X. Peng, S. Zhang, Y. Yu, and D. N. Metaxas, "Toward personalized modeling: Incremental and ensemble alignment for sequential faces in the wild," *Int. J. Comput. Vis.*, vol. 126, nos. 2–4, pp. 184–197, Apr. 2018.

[9] A. Jourabloo and X. Liu, "Pose-invariant face alignment via CNN-based dense 3D model fitting," *Int. J. Comput. Vis.*, vol. 124, no. 2, pp. 187–203, Sep. 2017.

[10] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3D solution," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 146–155.

[11] C. Gou, Y. Wu, F. Y. Wang, and Q. Ji, "Shape augmented regression for 3D face alignment," in *Proc. ECCV*, Amsterdam, The Netherlands, Jun. 2016, pp. 604–615.

[12] S. Z. Gilani, A. Mian, F. Shafait, and I. Reid, "Dense 3D face correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1584–1598, Jul. 2018.

[13] R. Anderson, B. Stenger, and R. Cipolla, "Using bounded diameter minimum spanning trees to build dense active appearance models," *Int. J. Comput. Vis.*, vol. 110, no. 1, pp. 48–57, Oct. 2014.

[14] S. Liao, D. G. Shen, and A. C. S. Chung, "A Markov random field groupwise registration framework for face recognition," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 36, no. 4, pp. 657–669, Apr. 2014.

[15] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models," in *Proc. BMVC*, Edinburgh, U.K., 2006, pp. 929–938.

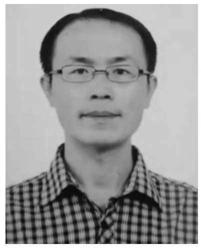
[16] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. CVPR*, Portland, OR, USA, Jun. 2013, pp. 1–8.

[17] R. Annunziata, C. Sagonas, and J. Cali, "Jointly aligning millions of images with deep penalised reconstruction congealing," in *Proc. ICCV*, Seoul, South Korea, Oct. 2019, pp. 81–90.

[18] S. Tulyakov, L. A. Jeni, J. F. Cohn, and N. Sebe, "Viewpoint-consistent 3D face alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2250–2264, Sep. 2018.

[19] Z. Q. Wei, Y. Yang, S. Zhang, and K. Yang, "Registration based on dual-feature Gaussian mixture model and dual-constraint spatial transformation," *J. Softw.*, vol. 29, no. 11, pp. 3575–3593, Nov. 2018.

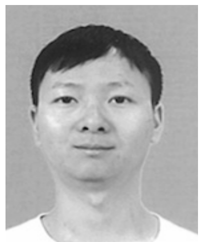
[20] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proc. ECCV*, Florence, Italy, 2012, pp. 679–692.



GANG ZHANG received the M.S. and Ph.D. degrees from the College of Information Science and Engineering, Northeastern University, China, in 2004 and 2009, respectively.

From 2011 to 2013, he was a Postdoctoral Research with Tsinghua University, China. Since 2013, he has been a Teacher and an Associate Professor with the Shenyang University of Technology, China. His research interests include machine learning, face recognition, image retrieval, and artificial intelligence.

HAN QIN is currently pursuing the master's degree with the Shenyang University of Technology, China. Her research interests include machine learning and face recognition.



YUDING KE is currently pursuing the master's degree with the Shenyang University of Technology, China. His research interests include machine learning and face recognition.

JIANSHENG CHEN received the M.S. degree from the School of Information Science and Technology, Tsinghua University, China, in 2002, and the Ph.D. degree from the Faculty of Engineering, The Chinese University of Hong Kong, China, in 2006.

From 2007 to 2009, he was a Postdoctoral Research with The Chinese University of Hong Kong. Since 2009, he has been a Teacher and an Associate Professor with Tsinghua University, China. His research interests include computer vision, and recognition and understanding of images and videos.

YANMIN GONG received the M.S. degree in information and communication engineering from Tsinghua University, China, in 2012, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Florida, USA, in 2016. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, The University of Texas at San Antonio, San Antonio, USA. Her research interests include machine learning, big data, the Internet of Things, cyber-physical systems, mobile computing, and wireless networks.