

Face Joint Alignment Using Local Method

ZHANG Gang^{1, a}, TANG Si Kan² and NIU Lian Qiang¹

¹Software School, Shenyang University of Technology, Shenyang 110023, China

²School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110023, China

^azhang_gang_1973@yahoo.com

Keywords: Face recognition, face alignment, joint alignment

Abstract. It is an under-determined problem that local methods are used for face alignment of an image, although good results can be obtained by using an auxiliary model or a priori information. In comparison, joint alignment using multiple face images of the same person has more advantages. In this paper, the rectangle round a face is acquired, and then logistic regressors are used to obtain the candidates of the regions round the landmark points. The non-parametric face shape models are used to constrain the configuration among the regions. On this basis, Generalized Procrustes Analysis is used for rigid joint alignment. Tests are carried out on LFW dataset, and the results show that joint alignment will cause overall drifting, and but helpful for facial landmark alignment in outer contour.

Introduction

Facial landmark points correspond to key positions in a face, which is of significance to expression analysis, face tracking, *etc.* Locating them has been an important research area in machine vision. Approaches for facial landmark location can be categorized into global approaches and local ones. Global approaches, *e.g.*, the approaches using active appearance models, regard facial landmark points as a whole. During the fitting of the models, the positions of these landmark points are updated until overall mean square error minimum is obtained. The approaches are susceptible to local distortion and local blur, and easy to fall into local extreme. Local approaches, *e.g.* the approaches using constrained local models, look upon landmark points as independent individuals. As such, the effect of local distortion and local blur is constrained into a small area. Besides, the approaches use facial shape models to constrain the configuration among the points. Thus it is ensured that the configuration conforms to knowledge of a person to faces. Certainly, the performance of local methods for face alignment is affected by sizes of face regions and features used ^[1]. Due to lack of information in a single face image, these problems are difficult to be solved.

When the subject of alignment is faces from a video or multiview, there will be more images to be available. Compared with face alignment for a single image, the use of multiple face images from the same person has more advantages, *e.g.* much information is available for the same point. These help to reduce the effect of local distortion and local blur further ^[2], and thus improve the alignment performance of the face images. Here we mainly want to figure out that joint alignment can improve alignment performance of a single face image or not.

Symbolic Normalization

For convenience to read, the symbols will be normalized. Let $X = \{x_1^1, \dots, x_i^j, \dots, x_M^N\}$ denote the set of facial landmark points obtained from multiple images by logistic regressors, where x_i^j describes the j -th landmark point of the i -th image, and M and N are the number of images and the number of facial landmark points in each image, respectively. Also, Let $Y = \{y_1^1, \dots, y_i^j, \dots, y_M^N\}$ denote the set of facial landmark points computed by joint alignment, where y_i^j represents the j -th landmark point of the i -th image. The meaning of symbols is graphically shown in Fig 1.

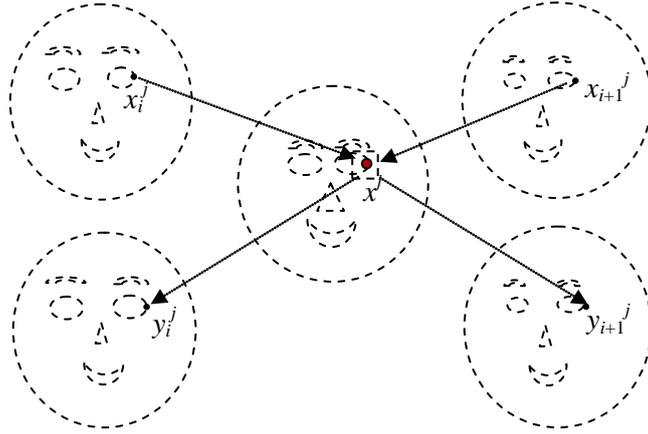


Fig.1 Description of symbol meaning

Joint Alignment Using Local Model

Here joint alignment method using local model will be denoted. For multiple face images of the same person, the rectangle round the face is obtained for each image. Then a local method for face alignment is used for rough location of facial landmark points. For location results from all images, Generalized Procrustes Analysis (GPA) [3] is used to perform rigid joint alignment. Then the results are mapped back to each face image, and thus joint alignment results can be obtained. The complete joint alignment method will be denoted as follows. It mainly contains three steps, *i.e.* rough location of landmark points, joint alignment of location results, and reverse mapping.

Joint alignment algorithm

Input: multiple face images of the same person

Output: joint alignment results

Steps:

- a) For each image, detect a face to obtain the rectangle round the face.
 - b) For each landmark point in a model, apply logistic regressors to locate region candidates.
 - c) Using face shape models to constrain the configuration among the regions.
 - d) Making use of GPA to perform rigid joint alignment.
 - e) Mapped back to original face images by means of transformation parameters.
-

Rough Location of Landmark Points

To avoid the case that part of a face is outside the rectangle around the face, the width and height of the rectangle are extended to 1.2 times original value, respectively. Then Logistic regressor [4] is used for the rectangle round the face to compute the response map of the j -th landmark point. Suppose that $I(x_i^j)$ is a candidate of the regions around the landmark point x_i^j , it will be zero mean and unit variance after the normalization. The logistic regressors can be described as

$$C(j, I) = w_i^T P(I(x_i^j)) + b_i \quad (1)$$

where I denotes a face image, the parameters w_i and b_i are gain and bias, respectively, and $P()$ is the function for region normalization.

The Gaussian distribution of the regions around the j -th landmark point can be obtained from a set of training samples. For each region, the normalization is carried out. Then a set of feature descriptors will be used to extract features. According to the feature vectors, the Gaussian distribution curves can be obtained and correspond to different feature descriptors.

Then the corresponding value of $C(j, I)$ can be obtained on these curves. These values are regarded as trust scores. Overall trust score can be computed by product-of-elements of these trust scores. For response map, the region of the maximal FS_j is used as location result. It needs to be noted that the set of points computed may not conform to knowledge of a person for a face. So face shape models are

usually applied. They are obtained from a set of training examples, and used to constrain the configuration among the points. To adapt to non-linear variations of faces, they are described non-parametrically. Supposed that $T = \{T_1^1, \dots, T_i^j, \dots, T_P^N\}$ is the set of face shape models where T_i^j denotes the j -th landmark point in the i -th model and P is the number of models, then the cost function can be used to measure rough location results and can be denoted as follows.

$$C_i = \min_{k \in [1, P]} \sum_{j=1}^N (x_i^j - T_k^j)^2 \quad (2)$$

After the above process is repeated for multiple face images, the corresponding sets of facial landmark points can be obtained, *i.e.* $X' = \{X_1', \dots, X_i', \dots, X_M'\}$ where X_i' corresponds to the i -th face image.

Joint alignment

Here GPA will be used for X' to perform rigid joint alignment. For any element in the set X' , *i.e.* X_i' , the centroid is computed, *i.e.* \bar{X}_i' , and then X_i' will be invariant to translation by $X_i' = X_i' - \bar{X}_i'$. The root mean square distance (RMSD) is computed for X_i' , *i.e.* S_i . The X_i' will be invariant to scale by $X_i' = X_i' / S_i$. Any element in X' , *e.g.* X_i' , is used as a reference, and the least squares method can be used to find the optimal overlapping angles between it and other elements, *i.e.* R_i . On this basis, the effect of rotation can be removed. Then all elements in the set X' are averaged, and the procrustes distance between the reference and the mean is computed. If the distance surpasses a threshold, the mean will be used as a new reference. The above process is iterated until the maximum number of iteration reaches or the procrustes distance is less than a threshold. As such, The sets of the facial landmark points can be obtained in the multiple face images, *i.e.* $X'' = \{X_1'', \dots, X_i'', \dots, X_M''\}$.

Reverse mapping

After X'' is obtained, it needs to be mapped back to original images. As such, three parameters needs to be considered, *i.e.* translation values, scale factors, and rotation angles. They are obtained from the joint alignment, *i.e.* \bar{X}_i' , S_i , and R_i . Reverse translation transforms the faces from a local coordinate system to a global coordinate system. Reverse scale recovers the difference of the sets of facial landmark points in scale. Reverse rotation considers the viewpoint in which the multiple face images are gathered. Reverse mapping can be denoted as follows.

$$Y_i = \bar{X}_i S_i R_i^{-1} (X_i'') + i' \quad (3)$$

It needs to be noted that S_i , R_i^{-1} and \bar{X}_i' are usually a combination of transformations.

Experiments

Tests is carried out in LFW dataset^[5] which contains more than 13000 color images of resolution 250×250 and some of the images contain two or more images of a face. The 500 pairs of face images are selected to form the dataset, *i.e.* two images for each individual. Two face images are regarded as a processing unit. AdaBoost classifiers are first used for detecting faces, and local binary pattern and gradient are used as feature descriptors at the stage of rough location. Then X'' and Y describe the alignment results obtained by local method and those obtained by joint alignment, respectively. Analyses will be carried out between them.

Measure

Suppose that $U_i = \{X_i' \text{ or } Y_i\}$ denote the alignment results of the i -th image, respectively, and GT_i is the corresponding ground truth, then similarity measure can be denoted as

$$JP = avg_{i \in [1, 2]} \left(\sum_{j=1}^N |U_i^j - GT_i^j|_2 \right) \quad (4)$$

where $avg()$ is the averaging function.

Test Results on Dataset

During testing, the overall shift which the joint alignment causes can be seen, as shown in Fig2. It sometimes reduces the alignment performance of a single face image to enhance the overall alignment results of two face images. However, an interesting phenomenon can also be found that the results obtained by the joint alignment migrate to one side of pose variations, which violates the tightness and but help to pull the landmark points in the outer contour to their actual positions. Besides, the measure above is used to measure X' and Y . However, it is found that X' excels Y on the whole. The case is mainly caused by overall drifting which rigid joint alignment causes.



Fig2. Overall shift which joint alignment causes (1) (1a), (2a) and (3a) correspond to results from local method (2) (1b), (2b) and (3b) are those from joint alignment

Conclusions

Joint alignment altogether with local method is used in the paper. Logistic regressors are used to obtain response maps of landmark points, and then trust scores are computed by means of distribution curves. On this basis, alignment results are obtained for a single face images. Then joint alignment is carried out for multiple face images of the same person. The results show that joint alignment will cause overall drifting which will influence performance of alignment, but helpful for facial landmark point location in the outer contour.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (61372176) and the Liaoning Province Education Department of China (L2015400).

References

- [1] J.M. Saragih, S. Lucey and J.F. Cohn: International Journal on Computer Vision Vol. 91 (2011), p. 200.
- [2] X.W. Zhao, X.J. Chai and S.G. Shan: European Conference on Computer Vision, edited by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid/ Springer Berlin Heidelberg, Florence (2012), p. 616.
- [3] J.C. Gower: Psychometrika Vol. 40 (1975), p. 33.
- [4] Y. Wang, S. Lucey and J.F. Cohn: Computer Vision and Pattern Recognition (2008), p. 1.
- [5] G.B. Huang, M. Ramesh, T. Berg and E. Learned-Miller, Amherst, Report (2007)