ISSN: 2664-9640

Research on Face Alignment Method Using Groupwise Registration

Yuding Ke[1], Ke Chen[2]

[1] Party School of the CPC Tianjin Jinnan District Committee, Tianjin 300350, China;

[2] School of telecommunication, Tianjin Railway Technical and Vocational College, Tianjin 300240, China.

Abstract

The traditional face registration method uses one instance face and one reference face to register. Different from the traditional registration method, face groupwise registration uses multiple face images to register at the same time. In this paper, we propose a combination of global rigid transformation and non-rigid distortion to improve the performance of face registration. Firstly, the initial position of the face calibration points in the group is determined by using the supervised descent method, and then the corresponding relationship between the face feature points in the group is found by using the RANSAC algorithm. Then, the non-rigid ICP algorithm is used to make the feature points in the reference face model non-rigidly distorted, so that each feature point in the reference face is transformed in the direction of the corresponding feature points of the instance face, and finally a local face registration method using the joint registration method is constructed. Experimental results on face database 300W show that the proposed method can improve the average location accuracy of face feature points compared with SDM.

Keywords

Face registration, groupwise registration, RANSAC algorithm, non-rigid ICP.

1. Introduction

As an important part of face recognition system, face registration connects the two key steps of face detection and feature matching, which has an important impact on the performance of face recognition. If the registration accuracy is high, then can improve the face recognition performance, and vice versa. Face registration methods can be divided into global methods and local methods. Global methods such as Active Appearance Model (AAM) [1] consider all the key points of the face at the same time, and all the key points are optimized simultaneously during the registration optimization period and are constrained by the global shape model of the face. Local methods such as Constrained Local Model (CLM) [2] consider the individual key points of the face, and each key point of the face can evolve separately. But during the registration optimization period, they are constrained by the shape model of the face, which weakens the correlation between the feature points. The results obtained by these methods are difficult to adapt to more complex real environment.

The face registration algorithm based on the regression model achieves ideal results in the limited environment and non-limited environment. In 2010, Dollar et al. [3] used the cascade shape regression method, which was the first representative regression-based method. By introducing shape index features, the influence of face pose changes is reduced to a large extent. In 2012, Dantone et al. [4] proposed a face registration algorithm based on regression forest, which achieved very ideal results on the LFW face database. In 2013, Xiong et al. [5] proposed a supervised descent method (SDM). By learning the descent direction and step size of the

function, it makes the function converge to the minimum quickly and avoids solving the Jacobian matrix and Hessian matrix. In 2014, Ren et al. [6] proposed LBF (local binary feature) algorithm, by calculating local binary features in the field of key feature points of face , each key point can generate random forest by self-learning, and finally do global linear regression based on cascade random forest.

In addition, the popular deep learning methods have been widely used in face registration in recent years. In 2014, Zhang et al. [7] used the convolutional neural network of cascade regression to locate facial feature points and optimize multiple targets at the same time, which further improved the accuracy of feature point location. In 2017, Kowalski et al. [8] proposed a new cascade deep alignment network (DAN), DAN added a key point heat map, input the whole face picture in each stage of training, in the state-of-the-art registration method at that time reduced the registration failure rate by 70%. In 2019, A novel end-to-end deep convolutional cascading (DeCaFA) architecture for face alignment is proposed [9], DeCaFA uses fully-convolutional stages to keep full spatial resolution throughout the cascade. In 2021, Chunze L al. [10] considers the interactions among facial landmarks and can be easily implemented on top of any convolutional backbone to boost the performance. The effectiveness of the proposed method is verified on WFLW, COFW and 300W datasets.

Different from the traditional face registration method which only uses one instance face, the method of groupwise registration uses multiple instances face to register the face. The groupwise registration method can model the information of each corresponding key point in the face, thus adding more constraints to each key point. Multiple face images can be from the same person or from different people in the groupwise registration method, which can avoid the failure of single image face registration. This is also the main reason for face registration in this paper.

2. Global Correspondence between Faces

2.1. Rigid Transformation of Facial Feature Points

In the registration of a group of faces, the global transformation method is used to transform the feature points. That is, a rigid geometric transformation model is used to transform the reference face into the group of instance faces. The homography transformation model is used to transform the face feature points to find an optimal homography matrix. For a given set of facial feature points, the best homography matrix between them can be found by randomly selecting feature points.

2.2. Correspondence between Feature Points

For the two face images, $I_1(x,y)$ is the reference face image and $I_2(x,y)$ is the instance face image. The goal is to calculate the optimal geometric transformation model between the reference face image $I_1(x,y)$ and the instance face image $I_2(x,y)$, that is, the homography matrix H. Any pair of feature points corresponding to the reference face $I_1(x,y)$ and the instance face $I_2(x,y)$ are denoted as $p=(x,y)$ and $p^{'}=(x^{'},y^{'})$, respectively, and they are converted into homogeneous coordinates and denoted as $p=(x,y,1)$ and $p^{'}=(x^{'},y^{'},1)$, respectively. In order to find the correspondence between them, it can be transformed into solving the optimal homography matrix between the reference face and the instance face. The specific steps are as follows:

(1) At least four pairs of feature points are randomly selected as the corresponding points of the initial hypothesis each time in the feature points set of the reference face and the instance

face, and then linear equations are established through the four pairs of feature points to calculate the homography matrix H, the transformation formula is shown in formula (1).

$$s\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

Where $\boldsymbol{s}$ is the scale coefficient, which usually $h_{33}=1$ to solve the matrix H. There are eight unknown quantities in matrix H. According to the basic principle of linear equations, at least four pairs of characteristic points are needed to solve the linear equations. In order to ensure the uniqueness of solution, it is necessary to ensure that all the four pairs of feature points are not collinear.

(2) Establish an error measurement function between the reference face and the instance face after the homography transformation, namely the cost function, which is used to measure the credibility between the selected feature points. The cost function is shown in formula (2).

$$\sum_{i=0}^{n}\left(x_i' - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2 + \left(y_i' - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}}\right)^2 \tag{2}$$

(3) The parameters of the homography transformation are calculated by repeating the above steps with four pairs of feature points randomly selected each time. Because of the uncertainty of the random selection of feature points, the number of iterations must be increased to improve the probability of obtaining the optimal geometric model. In the multiple iterations, select the transformation parameters to make the cost function value to the minimum.

3. Non-rigid Distortion between Faces within a Group

Due to the different shapes of each example face, there are local deformation in the face region, such as smile, mouth, frown and other ways will lead to local deformation of the face. The global rigid transformation model can not reflect the local deformation in the face region very well, so we consider the non-rigid distortion method to transform the facial feature points. Each feature point in the reference face is moved to make it move to the corresponding feature point in the instance face.

The transformation formula of a single feature point in the reference face by using the non-rigid distortion is shown in formula (3).

$$X_i^{new} = F(x_i) = \sum_{i=1}^{N} f(r) + a + bx_i + cy_i \tag{3}$$

Where, $x_i = (x_i, y_i)$ is the coordinates of the i-th feature point in the reference face, X_i^{new} is the i-th feature point on the deformed reference face, N is the number of feature points in the reference face, a, b, c represents affine transformation parameters, $f(r) = \exp(-k\|r\|)$ is a Gaussian kernel function [11], k takes $1/2\sigma^2$, the Gaussian kernel function can accelerate the registration process.

The reference face image is recorded as V, and the instance face image is recorded as P. Then, different affine transformations are performed on each feature point v_i in the reference face V, and finally the entire reference face is transformed. After the reference face V undergoes this elastic deformation, a new reference face is obtained as V', and the final goal is to make the obtained deformed reference face V' and the shape of the instance face P as close as possible. The individual movement of each feature point in the reference face may cause some feature points to fall into local minima or even result in registration failure. To avoid this situation, a

cost function can be constructed to constrain the evolution process of each feature point. The cost function formula is shown as formula (4).

$$E = \alpha E_d + \beta E_s + \gamma E_f \tag{4}$$

The cost function consists of three parts, namely E_d 、E_s 、E_f , which respectively represent the distance error of the data item, the error of the smoothing item and the registration error between the corresponding feature points. α , β , γ are weighting coefficients. E_d represents the similarity between the current deformed reference face and the corresponding feature points of the instance face, measured by the sum of squared Euclidean distances. The definition is shown in formula (5).

$$E_d = \sum_{i=1}^{N} \omega_i \left\| X_i v_i - D_i \right\|_2^2 \tag{5}$$

Where, v_i is the i-th feature point in the reference face V, and D_i is the feature point closest to v_i on the instance face *P*. ω_i is the error coefficient of the i-th feature point. If the feature point v_i in the reference face cannot find a corresponding point in the instance face *P*, that is, the closest point between v_i and P exceeds the set threshold, then the value of ω_i is set to 0, otherwise it is set to is 1.

E_s represents the consistency of the affine transformation parameters between adjacent feature points during the deformation process of the reference face feature points. The definition of E_s is shown in formula (6).

$$E_s = \sum_{\{(v_i, v_j) \in edge(V)\}} \left\| X_i - X_j \right\|^2 \tag{6}$$

Where, (v_i, v_j) is the line segment connected between v_i and v_j , and $edge(V)$ is the set of all edges in the reference face *V*.

E_f represents the distance between the deformed reference face and the corresponding feature points of the instance face. At the beginning of the iteration, the face feature point is marked as $K_l\,(1, 2, \cdots, 17)$, and its definition is shown in formula (7).

$$E_f = \sum_{i=1}^{17} \left\| X_{K_l} v_{K_l} - D_{K_l} \right\|^2 \tag{7}$$

The cost function is used to constrain the deformation process of the reference face feature points. The specific steps of fitting the reference face to the real face are as follows:

1) Initialize the affine transformation matrix $X_i^0\,(i = 1, 2, \cdots, N)$ of each feature point in the reference face, and set k=0;

2) For a given set of parameters α , β , γ , there are:

① For each feature point v_i^k on the reference face V^k, find its closest point on the instance face P as its corresponding point D_i^k ;

② Calculate the optimal affine transformation matrix $X_i^k\,(i = 1, 2, \cdots, N)$ for each feature point to minimize the value of the cost function E constructed above.

③ According to the obtained optimal affine transformation parameters to update the shape of the reference face, the update formula is (8).

$$V^{k+1} = \left\{ v_i^{k+1} \middle| i = 1, 2, \cdots, N \right\} = \left\{ X_i^k v_i^k \middle| i = 1, 2, \cdots, N \right\} \tag{8}$$

④Repeat the above ①-③ steps until the $\sum_{i=1}^{N}\left\|X_i^k - X_i^{k+1}\right\| < \varepsilon$ is satisfied;

Increase α and β, decrease γ, and repeat step 2) until the iteration stops when it converges.

4. Face Registration Method using Groupwise Registration

4.1. Algorithm Principles

Supervised descent method is used to determine the initial position of face calibration points. During groupwise registration, global rigid transformation and local non-rigid distortion are considered to improve the registration performance. During the iteration, the RANSAC algorithm [12] is used to find the global correspondence between the instance face and the reference face, and the rigid transformation is used to realize the global transformation between the faces, so as to determine the corresponding relationship between the feature points of the faces. non-rigid iterative closest point (NICP) [13] is used to consider the local distortion of the facial calibration points. An objective function is established to constrain the evolution of each feature point. The adjacent feature points are restricted by each other during the movement process. The affine consistency between adjacent feature points is maintained to achieve the movement of facial feature points to prevent feature points from getting into local minimum during the movement process.

4.2. Groupwise Registration Process

For a set of examples of face images, there are n face images in the set of face images, which are respectively recorded as I_1, I_2, I_3, ..., I_n. Unlike paired registration, which uses only one instance face, the proposed method uses a set of face images for registration. For each feature point in the instance face, more feature information can be used, so more constraints can be established. A face model is set up as the initial reference face in this set of face images. The main steps of the method are as follows.

(1) Initialize the position of facial feature points. For a given set of instance face images, we first need to determine the initial face shape for each instance face image, that is, the initial facial feature point location. The supervised descent method is used to locate the feature points of the group of examples. The obtained results are used as the initial location of the feature points of the group of examples.

(2) After determining the initial feature point location information of this set of examples, the generalized Procrustes analysis of these examples of faces is carried out to make the direction and size of the face images are consistent. Then, the average shape model is obtained by averaging the feature points of these instances. The principal component of the data is extracted by using the PCA algorithm, and the shape model of the instances is obtained.

(3) Each face in the set of instance face images is transformed into the shape model by twisting, and the instance face images are transformed into the shape model by using thin-plate spline function. At this time, the texture information of feature points is mapped, and an average texture model is obtained after averaging. The combination of face shape model and texture model is the face model, which is used as the reference face in the registration.

(4) Find the corresponding relation between the reference face and the feature points of this group of examples. The RANSAC algorithm is used to find the corresponding relationship between them, and four pairs of points are randomly selected as the initial corresponding points in the two sets of facial feature points at a time. The projection transformation matrix H is calculated and the cost function is established. After several iterations, the projection transformation parameters that minimize the cost function value are obtained.

(5) After determining the corresponding relationship between the feature points of the group, each feature point of the reference face can be moved independently. The local transformation

of the face is implemented by using the non-rigid ICP algorithm, which makes the feature points of the reference face move towards the feature points of the corresponding instance face, so as to achieve the optimal fitting of the reference face to each instance face.

(6) Update the reference face model. In step (5), a new reference face model is recalculated to replace the original reference face model. Then return to step (4) and repeat the steps (4)-(6) until the algorithm converges, stopping iteration. The flow chart of the registration algorithm is shown in Figure 1.

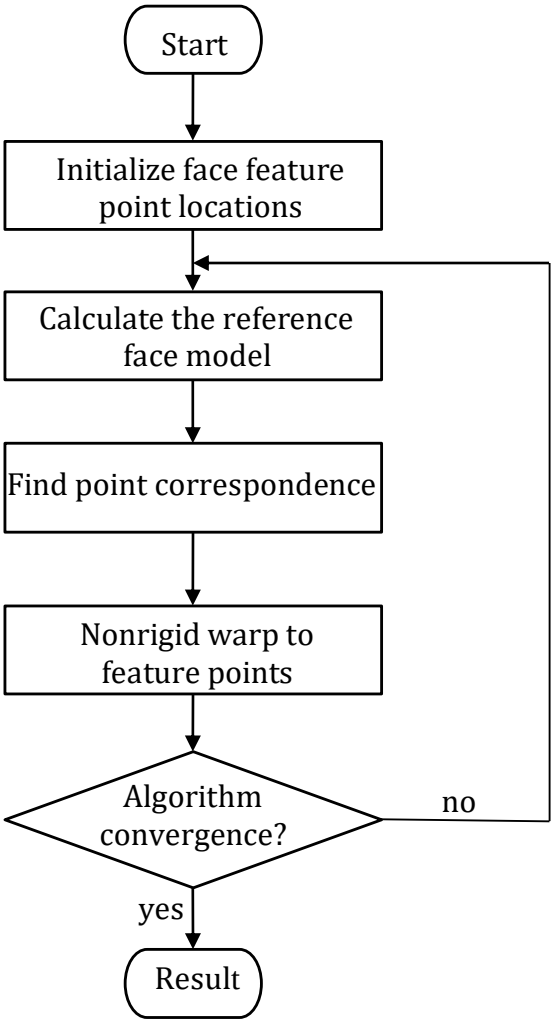

Figure 1: Groupwise registration algorithm flow chart

In the initial stage of registration, the value of feature point registration error E_f is very small, then we increase the constraint on the smooth term error E_s and relax the constraint on E_f, and the constraint on the data term error E_d is also from weak to strong. In step (6), set the maximum number of iterations, update the reference face model, calculate the data error $d^{'}$ of the reference face model before and after the update, and stop the iteration if $d^{'}$ is less than the previously set threshold τ.

4.3. Data Sets

The validation analysis was performed using the 300W database, which contains the LFPW [14], HELEN [15], AFW [16], and IBUG [17] datasets, and the face images from these datasets were re-labeled with face feature points, and each face image was labeled with 68 feature points.

ISSN: 2664-9640

4.4. Analysis of Experimental Results

The result of face registration obtained by SDM algorithm is used as the initial input data of the proposed face registration algorithm, which mainly verifies whether the accuracy of the proposed algorithm is improved compared with the supervised descent method.

Figure 2 and Figure 3 are the partial registration results of SDM algorithm and the proposed groupwise registration algorithm. In the first face, the registration results of the two methods are similar. In the second face, the SDM algorithm has a poor result in locating the eyebrow and eye region feature points in the face, which seriously deviates from the real location of feature points. However, the positioning effect of this method in the eye and eyebrow region feature points is significantly better than SDM algorithm. In the third face, the light of the face region is very weak, and the SDM algorithm is not accurate to locate the feature points of the chin region. However, the method in this paper makes the location results of the chin region closer to the real location. The fourth face has partial occlusion, the SDM algorithm in the mouth region and nose region of the feature points location is invalid, and the right cheek border location is also larger deviation. After the implementation of this method, the feature points of the nose, mouth and the facial outline region are all returned to the position which is very close to the real feature points. In order to accurately and quantitatively analyze the effect of the proposed registration algorithm and SDM algorithm, 10 face images are used as a group to test 10 groups of face, and the average error of the two methods is calculated.

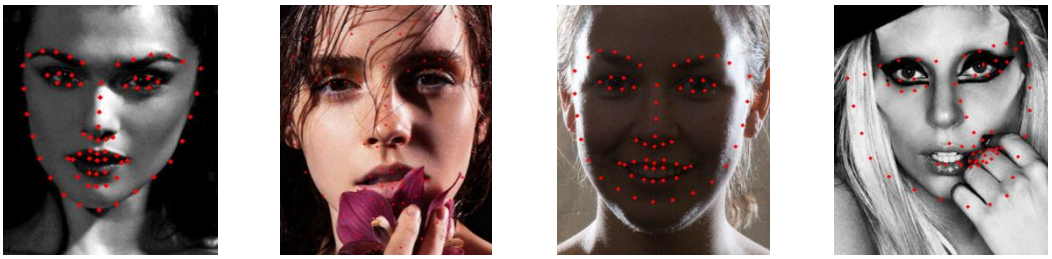

Figure 2: Supervised descent method partial registration effect chart

Figure 3: Partial registration effect diagram of the algorithm

Table 1 shows the average error comparison between the proposed registration algorithm and SDM algorithm on the face selected in the 300W database. From the table, we can see that except the second group of the algorithm precision is slightly worse than the SDM algorithm, the other groups of data show that the accuracy of the algorithm compared with SDM algorithm in this paper has improved to different degrees. In particular, the accuracy of the proposed method is improved more obviously than that of SDM algorithm in the six groups of 1,5,6,7,9 and 10 data.

Table 1: Accuracy comparison (mean error) of the algorithm and SDM algorithm

algorithm	1	2	3	4	5	6	7	8	9	10
SDM	9.16	6.89	10.20	9.35	11.72	13.43	11.83	9.48	11.81	7.17
Our method	6.85	7.07	10.10	8.99	9.56	8.27	7.73	8.36	9.34	5.71

Combined with the above analysis, SDM algorithm in the case of some facial feature points location results are invalid, after the registration method in this paper, the location of some regions of the face feature points in the original seriously deviated situation can return to the true position. We can see that the proposed method can improve the accuracy of face feature points registration to a certain extent compared with the SDM algorithm. At the same time, the proposed groupwise registration method is robust to illumination changes and occlusions.

5. Conclusion

In this paper, a new face registration method is proposed by combining rigid transform with non-rigid distortion, i.e., global rigid geometric transform is used to find the corresponding relationship between the feature points of the group. Then, the non-rigid transform is used to move each feature point of the reference face to the corresponding feature point of the instance face. This method takes the registration result of SDM algorithm as the initial feature points, and carries out experimental verification analysis on the commonly used face registration database 300W. Compared with the rigid face registration model, the non-rigid distortion makes the registration effect better. Compared with the registration results using SDM method, in the case of good initial registration average accuracy of SDM algorithm, the use of groupwise registration method of face registration in the positioning accuracy of facial feature points does not improve significantly, however, in the case of SDM algorithm for a group of facial feature points positioning average poor accuracy, this method in the group of facial feature points of the average positioning accuracy improvement is more obvious. In addition, for some SDM methods to register the face with local failure (such as mandible), this method can make the feature points of the face close to the true value. In general, in the face, the proposed registration method using multiple instances of face is superior to the SDM algorithm using only one instance of face.

This method is sensitive to the initial facial feature points. When the face pose is large, the performance of this method will decrease. The main reason is that the face pose change is not considered in the registration phase. In the next work, we will consider to improve the registration performance of the profile face, which makes the algorithm more universal, and at the same time it can handle the face images in different complex environments.

References

[1] T. F. Cootes, G. J. Edwards, C. J. Taylor: Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23 (2001) No.6, p.681-685.

[2] D. Cristinacce, T. F. Cootes: Feature detection and tracking with constrained local models, British Machine Vision Conference (Edinburgh, UK, September 4-7, 2006). Vol.1, p.929-938.

[3] P. Dollár, P. Welinder, P. Perona: Cascaded pose regression, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (San Francisco, USA, June 13-18, 2010). Vol.1, p. 1078-1085.

[4] M. Dantone, J. Gall, G. Fanelli: Real-time facial feature detection using conditional regression forests, IEEE Conference on Computer Vision and Pattern Recognition (Malatya, Turkey, May 16-19, 2012). Vol.1, p. 2578-2585.

[5] X. Xiong, F. Torre: Supervised descent method and its applications to face alignment, IEEE Conference on Computer Vision and Pattern Recognition (Portland, USA, June 23-28, 2013). Vol.1, p. 532-539.

[6] S. Ren, X. Cao, Y. Wei: Face alignment at 3000 fps via regressing local binary features, IEEE Conference on Computer Vision and Pattern Recognition (Columbus, USA, June 23-28, 2014). Vol.1, p.1685-1692.

[7] Z. Zhang, P. Luo, C. C. Loy: Facial landmark detection by deep multi-task learning, European Conference on Computer Vision (Switzerland, Switzerland, 2014). Vol.1, p.94-108.

[8] M. Kowalski, J. Naruniec, T. Trzcinski: Deep alignment network: A convolutional neural network for robust face alignment, IEEE Conference on Computer Vision and Pattern Recognition Workshops (Honolulu, USA, July 21-26,2017). Vol.1, p.2034-2043.

[9] A. Dapogny, M. Cord and K. Bailly: DeCaFA: Deep convolutional cascade for face alignment in the wild, International Conference on Computer Vision (Seoul, Korea, October, 27, 2019). Vol.1, p.6892-6900.

[10] C. Lin, B. Q. Wang: Structure-Coherent deep feature learning for robust face alignment, IEEE Transactions on Image Processing, vol.30 (2021), No.1, p.5313-5326.

[11] D. Ramanan, X. Zhu: Face detection, pose estimation, and landmark localization in the wild, IEEE Conference on Computer Vision and Pattern Recognition (Providence, USA, June 16-21,2012). Vol.1, p.2879-2886.

[12] M. A. Fischler, R. C. Bolles: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, Vol.24 (1981), No.6, p.381-395.

[13] B. Amberg, S. Romdhani, T. Vetter: Optimal step nonrigid icp algorithms for surface registration, IEEE Conference on Computer Vision and Pattern Recognition (Minneapolis, USA, June 17-22, 2007). Vol.1, p.1-8.

[14] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman: Localizing parts of faces using a consensus of exemplars, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.35 (2013), No.12, p.2930-2940.

[15] V. Le, J. Brandt, Z. Lin: Interactive facial feature localization, European Conference on Computer Vision (Florence, Italy, October 7-13, 2012). Vol.1, p.679-692.

[16] D. Ramanan, X. Zhu: Face detection, pose estimation, and landmark localization in the wild, IEEE Conference on Computer Vision and Pattern Recognition (Providence, USA, June 16-21, 2012). Vol.1, p.2879-2886.

[17] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou: 300 faces in-the-wild challenge: The first facial landmark localization challenge, IEEE International Conference on Computer Vision Workshops (Sydney, Australia, December 2-8, 2013). Vol.1, p.397-403.