2012 International Workshop on Nonlinear Circuits,
Communications and Signal Processing
NCSP'12, Honolulu, Hawaii, USA. March 4-6, 2012

Nonrigid Face Alignment Using Relative Motion Prior

Yang Yang†, Yuehu Liu†, Yuanqi Su†, Yoshifumi Nishio‡

†The Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University, Xi'an 710049, China
Email: yyang@mail.xjtu.edu.cn

‡The Department of Electrical and Electronic Engineering
Tokushima University, Tokushima 770-8506, Japan
Email: nishio@ee.tokushima-u.ac.jp

Abstract

We propose a novel method to automatically align face images using a single template. This is a challenge work since the appearance variance caused by various expressions. The core of our method is to utilize the relative motion prior for constraining nonrigid deformation parameters. The proposed method is independent of the appearance model and available for unseen face images. Considering the computation efficiency, our model fitting is solved in the inverse compositional framework. Experiments on face images and videos demonstrate the effectiveness of the proposed method.

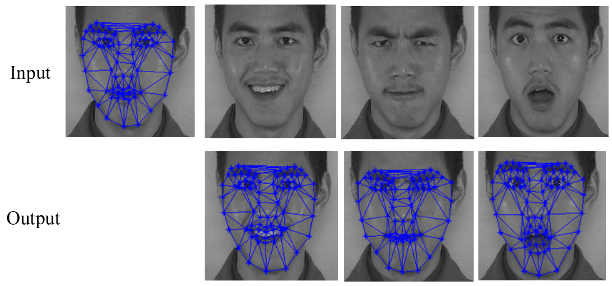


Figure 1: Illustration of the nonrigid face alignment problem. We take a single face image with neutral expression as the template, to align other expression faces automatically.

1. Introduction

Face image alignment is an important topic in the computer vision and image processing. It has various applications in the feature detection, motion analysis and tracking. There are two main categories of alignment algorithms. The first category is based on the generative model, like the Active Appearance Model (AAM)[1][2], which achieves accurate fitting results and high efficiency using the analysis-by-synthesis strategy. However, as shown in the previous work[3], it falls short of successfully aligning unseen faces, since the model generalization capability is limited in training examples. Another category is based on the template fitting[4][5]. Without the appearance model, template-based method is a kind of image-to-image face alignment. To capture the appearance variation, this type of methods also needs multiple templates.

In this paper, we propose a novel face alignment method using a single template. The goal is to fit the only neutral expression appearance to other expression faces. Figure 1 presents the illustration of the nonrigid face alignment task. The proposed method can also be used for the automatical landmark labeling, which reduces the labor-intensive work for preparing training examples and removes labeling inconsistencies among different labelers. Furthermore, it can be applied to track face expressions in video sequences.

Same as other template-based methods, this single template alignment suffers from appearance variations across various expressions. Instead of handling the face appearance model, we address the above problem by constraining geometric parameters. Motivated by the relative motion model being successfully utilized for the expression synthesis[6] and facial animation[7] work, we add the relative motion prior (RMP) to the face alignment algorithm. Since the RMP is learned for common face motion regulations, it has two benefits for the face alignment: 1). local fitting information will be propagated to the entire face, which helps escaping from the local minima; 2). the fitting is more robust with the prior of warping parameters, since correct facial motion constraints reduce the negative fitting caused by appearance variations.

Specifically, we tackle the optimization problem of face alignment by introducing a regularization term about the RMP. Taking the computational cost into account, we solve the model fitting under the inverse compositional (IC) framework. The most time-consuming parameters are precomputed off-line. While in the on-line process, geometric parameters are updated efficiently. Comparison results tested on different unseen faces demonstrate the proposed method achieves better performances than the conventional alignment method in term of the fitting accuracy.

The remainder of this paper is organized as follows. In Section 2, we give a brief introduction of the image alignment problem and state the outline about model fitting: the standard IC algorithm. Section 3 presents the proposed face alignment method and the details about its model fitting. In Section 4, the proposed method is tested by the comparison experiments. Finally, conclusions are given in the last section.

2. Conventional Image Alignment and the IC Algorithm

The conventional image alignment problem is presented in the Lucas-Kanade algorithm[4]. Let T and I denote a template image and a target image, respectively. The goal is to find warp parameters p that minimize the cost function of the form

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

where $\mathbf{W}(\mathbf{x};\mathbf{p})$ is the warping function for the pixel coordinates $\mathbf{x}$ in the fitting area. $\mathbf{W}(\mathbf{x};\mathbf{p})$ is often chosen as the piecewise affine warp to model nonrigid transformation. We apply the point distribution model (PDM) to present the face shape:

$$\mathbf{S} = \bar{\mathbf{S}} + \mathbf{Q}\mathbf{p}, \tag{2}$$

where $\bar{\mathbf{S}}$ and $\mathbf{Q}$ are the mean and basis of variation respectively.

The inverse compositional (IC) algorithm[4] is proposed to solve this optimization problem efficiently. The key idea is to change the role of T and I when updating $\triangle\mathbf{p}$:

$$E(\triangle\mathbf{p}) = \sum_{x} \|T(\mathbf{W}(\mathbf{x};\triangle\mathbf{p})) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))\|^2. \tag{3}$$

And then the incremental warp $\mathbf{W}(\mathbf{x};\triangle\mathbf{p})$ is inverted and composed with the current estimate to the warping function iteratively:

$$\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\triangle\mathbf{p})^{-1}. \tag{4}$$

This change enables that the most time-consuming parameters can be estimated off-line.

3. Nonrigid Face Alignment with the RMP

3.1. Proposed Model

We tackle the expression face alignment problem using a single template by introducing a regularization term about the RMP. In this section, we first give the RMP learning process.

The relative motion vector is defined as

$$\mathbf{S}_r = \mathbf{S} - \mathbf{S}_0, \tag{5}$$

where $\mathbf{S}_0$ denotes the face with neutral expression. The relative motion vector is modeled by the principal components analysis (PCA) and projected into a low-dimensional space. $\mathbf{S}_r$ is parameterized by $\mathbf{p}_r$

$$\mathbf{S}_r = \bar{\mathbf{S}}_r + \mathbf{Q}_r\mathbf{p}_r, \tag{6}$$

where $\mathbf{Q}_r$ contains unit eigenvectors of the projection matrix and $\bar{\mathbf{S}}_r$ is the mean value. Figure 2 shows the effect of varying the first three component's parameters respectively. Now the RMP can be estimated from the anisotropic Gaussians distributions in the PCA model.

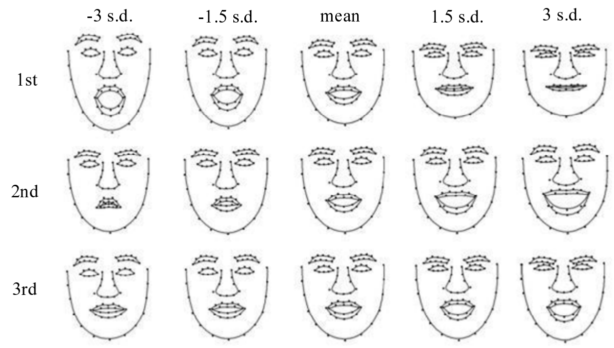


Figure 2: The effects of varying first three component parameters in the relative motion model. We add the reconstructed motion vector to a mean neutral face. Each component's parameter is shown in a row with variation between ±3 s.d.

The optimization problem with the RMP is formulated to minimize following energy function:

$$E_c(\mathbf{p}) = \sum_{\mathbf{x}} \|T(\mathbf{x}) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))\|^2 + \lambda_c R(\mathbf{p})^T \Sigma_r^{-1} R(\mathbf{p}),$$

where the regulation term $R(\mathbf{p})$ is the relative motion vector calculated by the current face shape parameter $\mathbf{p}$, by combining Eqs. (2), (5) and (6),

$$R(\mathbf{p}) = \mathbf{Q}_r^{-1}(\bar{\mathbf{S}} + \mathbf{Q}\mathbf{p} - \mathbf{S}_0 - \bar{\mathbf{S}}_r). \tag{7}$$

And Σ_r is the variance matrix $diag(\sigma_1^2, \sigma_2^2, ..., \sigma_N^2)$. The variance σ_i^2 equals to the ith eigenvalue. λ_c is a regularization coefficient.

3.2. Model fitting

The model fitting is to solve for optimal increments to parameters $\mathbf{p}$. As an extension of the IC algorithm, we adopt the inverse warping technique to minimize $E_c(\mathbf{p})$. The difference is that, in each iteration, the inverse incremental parameters should be reflected back to the regularization term.

$$E_c(\triangle\mathbf{p}) = \sum_{x} \|T(\mathbf{W}(\mathbf{x};\triangle\mathbf{p})) - I(\mathbf{W}(\mathbf{x};\mathbf{p}))\|^2 \tag{8}$$
$$+\lambda_c R(\mathbf{p} + \frac{\partial\mathbf{p}'}{\partial\triangle\mathbf{p}}\triangle\mathbf{p})^T \Sigma_r^{-1} R(\mathbf{p} + \frac{\partial\mathbf{p}'}{\partial\triangle\mathbf{p}}\triangle\mathbf{p}),$$

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

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

The function $E_c(\triangle\mathbf{p})$ is nonlinear with respect to $\triangle\mathbf{p}$, which can be linearized by taking the first order Taylor expansion. Let $\mathbf{A} = \bigtriangledown T \frac{\partial \mathbf{W}}{\partial \mathbf{p}}$ and $\mathbf{B} = \frac{\partial \mathbf{R}}{\partial \mathbf{p}} \frac{\partial \mathbf{p}'}{\partial \triangle \mathbf{p}}$.

$$E_c(\triangle \mathbf{p}) \approx \sum_x \|T + \mathbf{A}\triangle \mathbf{p} - I(\mathbf{W}(\mathbf{x};\mathbf{p}))\|^2 \tag{10}$$
$$+\lambda_c \left[R(\mathbf{p}) + \mathbf{B}\triangle \mathbf{p}\right]^T \Sigma_r^{-1} \left[R(\mathbf{p}) + \mathbf{B}\triangle \mathbf{p}\right].$$

The Hessian matrix is

$$\mathbf{H} = \sum_x \mathbf{A}^T\mathbf{A} + \lambda_c \mathbf{B}^T \Sigma_r^{-1} \mathbf{B}.$$

Hence, the solution of the above equation can be stated as:

$$\triangle \mathbf{p} = \mathbf{H}^{-1} \left(\sum_x \mathbf{A}^T (I(\mathbf{W}(\mathbf{x};\mathbf{p}) - T) - \lambda_c \mathbf{B}^T \Sigma_r^{-1} R(\mathbf{p}) \right).$$

According to Eq. (7), we have $\frac{\partial R}{\partial \mathbf{p}} = \mathbf{Q}_r^{-1}\mathbf{Q}$. The main remains to be described for this fitting problem is how to compute $\frac{\partial \mathbf{p}'}{\partial \triangle \mathbf{p}}$ of $\mathbf{B}$.

We set $\mathbf{S}'$ as the reconstructed face shape from $\mathbf{p}'$ according to Eq. (2). Combining Eq. (9), it is computed as $\mathbf{S}' = \mathbf{W}(\mathbf{S}_0 - \mathbf{Q}\triangle \mathbf{p}; \mathbf{p})$. Applying the chain rule,

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

We choose the piecewise affine warp for $\mathbf{W}(\mathbf{x};\mathbf{p})$. Given two corresponding triangles in the base mesh and target mesh, a set of affine transform parameters $(a_i, i = 1, ..., 6)$ of this warp can be computed in the closed form. For each pixel in this base mesh, which is represented by coordinates $(x, y)^T$, its location after the piecewise affine warp is required as:

$$\mathbf{W}(\mathbf{x};\mathbf{p}) = (a_1 \cdot x + a_2 \cdot y + a_3, a_4 \cdot x + a_5 \cdot y + a_6)^T.$$

Therefore, $\frac{\partial \mathbf{W}(\mathbf{x};\mathbf{p})}{\partial x} = (a_1, a_4)^T$ and $\frac{\partial \mathbf{W}(\mathbf{x};\mathbf{p})}{\partial y} = (a_2, a_5)^T$.

To compute the above derivative of warp function, the problem is to decide which triangle do we use for each vertex, since each vertex may connect several triangles. Our strategy is to average the affine parameters of every triangle that shares the vertex. This way achieves a smooth warp.

The parameter $\mathbf{B}$ must be recomputed in each iteration, so we have analyzed the computational cost of the proposed method. Let n and K denote the dimension of $\mathbf{S}$ and $\mathbf{p}_r$, respectively, and N denote the number of image vectors. The computational cost of the conventional IC algorithm is $O(nN + n^2K)$. The extra computations in the per-iteration of proposed method are the computing $\mathbf{B}$: $O(n^2K)$ and adding $\mathbf{B}$ into the Hessian: $O(n^2K)$. However, since $(K, n \ll N)$, the extra computational cost is negligible.

4. Experiments

We select 300 face images from 80 subjects in the Cohn-Kanade Facial Expression Database[8]. 80 specific landmarks are labeled around facial components. The template size of face region is 126×126 pixels. To accommodate global lighting variation, the intensity of face region is normalized in advance. In our experiments, the point distribution model and the relative motion model are both remained 95% principle components and the regularization coefficient λ_c is set to 100 .

Since good initial parameters are crucial to the success of face alignment, there have been many robust methods[2] for estimating the initialization. However, our main purpose is to discuss the effect of combining the RMP. We simply locate the initial face shape using the detected eye corners to target images. For comparison purpose, same initial parameters are employed in different alignment methods.

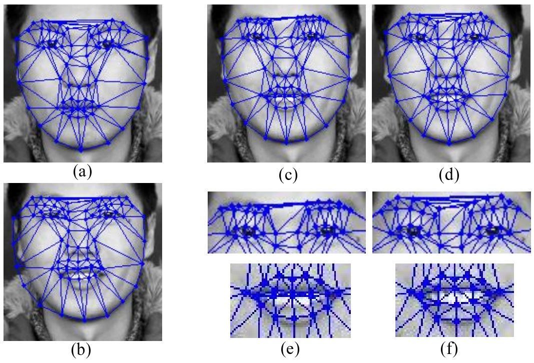
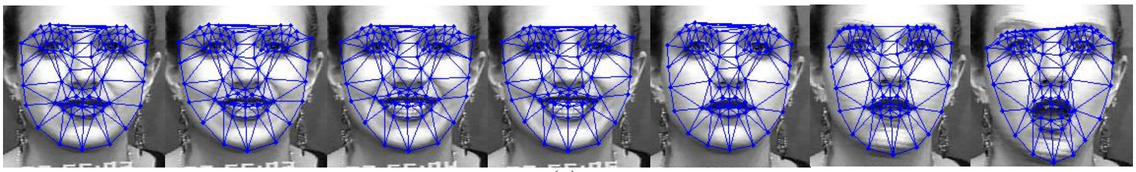


Figure 3: Comparison of face alignment results. (a) The single template with neutral expression. (b) The test image with the initial position and shape. (c) and (d) are the alignment results from the conventional alignment algorithm and our proposed method, respectively. More details on close-up views of the eye and mouth areas are shown in (e-f) for each method.

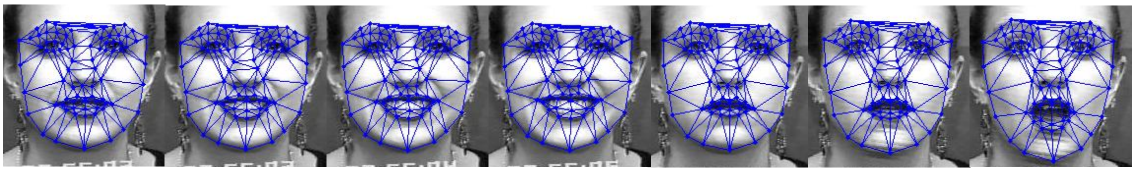
Figure 3 shows some examples of alignment results. The number of iterations for the model fitting is 10. We can see that our method performs better than the conventional alignment introduced in Sec. 2. Notice the details shown in Fig. 3(e-f), there is a large appearance difference: the teeth, but with the RMP, the deformed mouth shape is correctly estimated.

Tracking facial features from video sequences is another application of the proposed method. The first frame is often supposed as the alignment template. Figure 4 shows some examples of alignment results. We can see the proposed method achieves much better performance. For the smile expression, the mouth shape is deformed with a natural looking. For the surprise expression, with the constraints of the RMP, one point motion of the eyebrow is propagated to others, so that it avoids some local minimum situations.

Different from the above experiments, here we would like to quantitatively evaluate the alignment performance. To fit expression faces with a large appearance variance has more challenges. Therefore, we specially employ a testing database



(a)



(b)

Figure 4: Alignment results for video sequences by the conventional alignment algorithm (a) and proposed method (b).

including 70 images with smile and surprise expressions. Comparison results are demonstrated by computing the MSE between fitting results and the ground truth. Mouth is a difficulty in the fitting process. We additionally give the evaluation results of the mouth alignment. The proposed method gets better fitting accuracy than the conventional alignment method on both entire faces and local mouth fittings (Fig. 5).

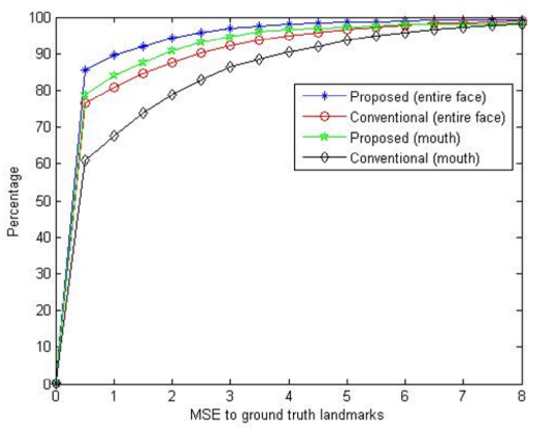


Figure 5: Alignment performance comparisons between the proposed method and the conventional method. We plot the Cumulative Density Function of the MSE on both entire faces and local mouth fittings.

5. Conclusions

In this paper, a nonrigid expression face alignment method is proposed, using a single template of the target neutral face. We employ the relative motion prior (RMP) as the constraint of warp parameters. The model fitting can be efficiently solved in the inverse compositional framework. Compared to the conventional alignment method, the proposed method outperforms on the single template alignment.

Acknowledgements: This work was supported by the Fundamental Research Funds for the Central Universities and by the National Natural Science Foundation of China (90920008, 91120009).

References

[1] T. Cootes, G. Edwards, and C. Taylor, “Active Appearance Models,” *IEEE Trans. Pattern Analysis and Machine Inteligence*, vol. 23, no. 6, pp. 681-685, 2001.

[2] M. Zhou, L. Liang, J. Sun, and Y. Wang, “AAM based Face Tracking with Temporal Matching and Face Segmentation,” *IEEE Conf. on Computer Vison and Pattern Recognition*, pp. 701-708, 2010.

[3] R. Gross, I. Matthews, and S. Baker, “Generic vs. Person Specific Active Appearance Models,” *Image and Vision Computing*, vol. 23, no. 11, pp. 1080-1093, 2005.

[4] S.Baker and I. Matthews, “Lucas-Kanada 20 Years on: A Unifying Framework,” *Int’l J. Computer Vision*, vol. 56, no. 3, pp. 221-255, 2004.

[5] J. Zhu, L. Gool, and S. Hoi, “Unsupervised Face Alignment by Robust Nonrigid Mapping,” *IEEE Conf. on Computer Vision*, pp. 1265-1272, 2009.

[6] Y. Du and X. Lin, “Emotional Facial Expression Model Building,” *Pattern Recognition Letters*, vol. 24, no. 16, pp. 2923-2934, 2003.

[7] Y. Yang, N. Zheng, et al., “Interactive Facial Sketch Expression Generation Using Local Constraints,” *IEEE Conf. on Intelligent Computing and Intelligent Systems*, vol. 2, pp. 864-868, 2009.

[8] T. Kanade, J.F. Cohn, and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” *Automatic Face and Gesture Recognition*, pp. 46-53, 2000.