Multi-resolution 3D Morphable Models and Its Matching Method

Bong-Nam Kang, Hyeran Byun
Department of Computer Science, Yonsei University, Korea
{bongnam, hrbyun}@yonsei.ac.kr

Daijin Kim
Department of Computer Science and Engineering, POSTECH, Korea
dkim@postech.ac.kr

Abstract

The inverse compositional image alignment (ICIA) is known as an efficient matching method for 3D morphable models (3DMMs). However, it requires a long computation time since the 3D face models consist of a large number of vertices. Also, it requires to recompute the Hessian matrix using the visible vertices every iteration. For a fast and an efficient matching, we propose the efficient and accurate hierarchical ICIA (HICIA) matching method for 3DMMs. The proposed matching method requires multi-resolution 3D face models and the Gaussian image pyramid. The multi-resolution 3D face models are built by sub-sampling at the 2:1 sampling rate to construct the lower-resolution 3D face models. For more accurate matching, we use a two-stage model parameter update that only updates the rigid and the texture parameters and then updates all parameters after the initial convergence. We present several experimental results to prove that the proposed method shows better performance than that of the conventional ICIA matching method.

1 Introduction

Recently, many researchers have been interested in the human face analysis To conduct research in these topics, we need to perform the face modeling and matching method of the model to the input image. There are two kinds of face modelings: the 2D face modeling and the 3D face modeling.

In the 2D-based face modeling, there are many approaches. Active contour models (ACMs) are only used to extract the facial contour of frontal-view faces and the performance crucially depends on the weight for which manually parameters tuning may be inevitable in many applications [8]. In active shape models (ASMs), the linear characteristics of ASMs limit their application to their small range of shape variations [4]. Active appearance models (AAMs) have the efficient matching methods such as the inverse compositional simultaneous update [6] and the inverse compositional project out [9]. They do not fulfill the two requirements such as the stability and the operating speed at the same time. The 2D-based methods have still the limitations that the matching methods are unstable and inaccurate for representing the input facial images.

To overcome these limitations, many researchers have proposed the 3D-based methods whose shape or texture can be controlled by a compact set of parameters [2, 5, 7]. For example, the 3DMMs use the most detailed 3D shape consisting of thousands of vertices. In the 3DMMs, there are two famous matching methods such as a stochastic Newton optimization (SNO) [3] and an inverse compositional image alignment (ICIA) [10]. While more accurate, the SNO requires a huge amount of computation time because it is necessary to recompute the Jacobian and the Hessian matrices at every iteration. On the other hand, the ICIA is an efficient method because the derivatives are pre-computed once for all iterations. Despite of this merits, the ICIA took about 30 seconds for fitting the model to a image [10].

To reduce the computation time of the conventional ICIA matching method, we propose a fast and an efficient hierarchical ICIA matching method for 3DMMs that guarantees the more efficient and accurate fitting performance. It requires a set of multi-resolution 3D face models and the matching has been performed hierarchically from the low resolution matching to the high resolution matching. For the more accurate matching, we take the two-stage model parameter update where the rigid and texture parameters are updated at the initial matching and all parameters are updated after the

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

convergence of the initial matching.

This paper is organized as follows: Section 2 describes the conventional ICIA matching method. In Section 3, we present the proposed method to reduce the computation time of matching. Section 4 describes experimental results. Finally, Section 5 draws a conclusion.

2 3D ICIA Matching Method

The matching method performs the matching of the 3DMM to a 2D face image and seeks the parameters of the 3DMM that express the model texture as close to the input image as possible. In the 3D ICIA matching method, the cost function C is an iteratively minimized log-likelihood that is the sum of the squares of the difference texture between the model and the image defined in the reference frame $\mathbf{u} = (u, v)^T$ as:

$$\sum_{\Omega(\boldsymbol{\alpha}^d, \boldsymbol{\rho}^d)} \big[\mathbf{t}(\mathbf{p}^{-1}(\mathbf{p}(\mathbf{u}_i; \boldsymbol{\alpha}^d + \delta\boldsymbol{\alpha}, \boldsymbol{\rho}^d + \delta\boldsymbol{\rho}); \boldsymbol{\alpha}^d, \boldsymbol{\rho}^d); \delta\boldsymbol{\beta}) - \mathbf{t}^{-1}(\mathbf{I}(\mathbf{p}(\mathbf{u}; \boldsymbol{\alpha}^t, \boldsymbol{\rho}^t)); \boldsymbol{\beta}^t)\big]^2, \quad (1)$$

where the superscripted parameters by d refer to the parameters at which the derivatives are computed, the parameters $\boldsymbol{\alpha}^t$, $\boldsymbol{\rho}^t$, and $\boldsymbol{\beta}^t$ are the current shape, motion, and texture parameters, respectively. $\Omega(\boldsymbol{\alpha}^d, \boldsymbol{\rho}^d)$ is a set of the visible vertices. This method was further detailed in [10].

3 The Proposed Hierarchical ICIA Matching Method

Although the ICIA matching method is efficient, it requires a lot of computation time($\approx$ 30 seconds) for the 3DMM matching. The time complexity of the conventional ICIA matching method is $O((N_s+N_t)\cdot N_{vv}+N_{vv}\log N_{vv} + (N_s+N_t)^2 \cdot N_{vv})$, where N_s, N_t, and N_{vv} are the dimension of the shape model, the dimension of the texture model and the number of the visible vertices, respectively. The first terms are the update of the shape and texture parameters, the second terms are the computation of the median of absolute deviations which is necessary for sorting the residuals to designate the outliers, and the third terms are due to the re-computation of the Hessian matrix. As shown by, we need to reduce the number of vertices as many as possible to reduce the matching time. To meet this requirement, we propose the hierarchical ICIA matching method.

3.1 Multi-Resolution 3D Face Models

The multi-resolution 3D face models are constructed by sub-sampling the shape model $\mathbf{S}$ that is formed by the mean shape and the shape basis vectors, formed by the mean texture and the texture basis vectors. The shape model $\mathbf{S}_{level3}$ and texture model $\mathbf{T}_{level3}$ at the level3 are just the shape model $\mathbf{S}$ and the texture model $\mathbf{T}$, respectively. Next, the shape and texture at the level3 are sub-sampled at the 2:1 sampling rate to construct the shape model $\mathbf{S}_{level2}$ and texture model $\mathbf{T}_{level2}$. Finally, the shape model and texture model at the level2 are sub-sampled at the 2:1 sampling rate to construct the shape model $\mathbf{S}_{level1}$ and texture model $\mathbf{T}_{level1}$.

3.2 Hierarchical ICIA Matching Method

After generating the multi-resolution 3D face models, we apply the proposed hierarchical ICIA matching method. Before the multi-resolution 3D face models to the input image, we generate the Gaussian image pyramid (GIP), which is a hierarchy of the low-pass filtered version of the original image such that the successive level corresponds to the lower frequency image, where the low-pass filtering is conducted using the convolution with a Gaussian filter kernel. At the initialization, the ICIA matching method requires the correspondences between some of the model vertices and the input face image manually [10]. But the proposed matching method tries an automatic initialization by aligning the pre-designated eye's positions in the face model and the detected eye's positions by the face detector. For updates of the model parameters, we use two-stage model parameter update (TSPU) that only updates the rigid and texture parameters at initial matching, and then updates all model parameters.

To summarize, the proposed hierarchical ICIA matching method takes the following steps:

(1) Generate the GIP from the input image $\mathbf{I}$:
$\mathbf{I}_{level1}, \mathbf{I}_{level2}, \mathbf{I}_{level3}$.
(2) Detect face and eyes using the face detector.
(3) Set the initial parameters for the first layer ICIA matching: $\boldsymbol{\alpha}^0 = \mathbf{0}, \boldsymbol{\beta}^0 = \mathbf{0}, \boldsymbol{\rho} = Face - detector$.
(4) Perform the first layer ICIA fitting:
$\mathbf{ICIA\_matching}(\mathbf{I}_{level1}, \mathbf{S}_{level1}, \mathbf{T}_{level1}, \boldsymbol{\alpha}^0, \boldsymbol{\rho}^0, \boldsymbol{\beta}^0)$.
(5) Set the initial parameters for the second layer ICIA matching: $\boldsymbol{\alpha}^0 = \boldsymbol{\alpha}^*_{level1}, \boldsymbol{\rho}^0 = \boldsymbol{\rho}^*_{level1}, \boldsymbol{\beta}^0 = \boldsymbol{\beta}^*_{level1}$
(6) Perform the second layer ICIA fitting:
$\mathbf{ICIA\_matching}(\mathbf{I}_{level2}, \mathbf{S}_{level2}, \mathbf{T}_{level2}, \boldsymbol{\alpha}^0, \boldsymbol{\rho}^0, \boldsymbol{\beta}^0)$.
(7) Set the initial parameters for the third layer ICIA matching: $\boldsymbol{\alpha}^0 = \boldsymbol{\alpha}^*_{level2}, \boldsymbol{\rho}^0 = \boldsymbol{\rho}^*_{level2}, \boldsymbol{\beta}^0 = \boldsymbol{\beta}^*_{level2}$
(8) Perform the third layer ICIA mathcing:
$\mathbf{ICIA\_matching}(\mathbf{I}_{level3}, \mathbf{S}_{level3}, \mathbf{T}_{level3}, \boldsymbol{\alpha}^0, \boldsymbol{\rho}^0, \boldsymbol{\beta}^0)$.

4 Experimental Results

To show the validity of the proposed method, we define some performance measures such as the normalized correlation and the root mean squared error (RMSE). First, we define the normalized correlation as:

$$C = \frac{\boldsymbol{\alpha}^T \cdot \tilde{\boldsymbol{\alpha}}}{\|\boldsymbol{\alpha}\| \cdot \|\tilde{\boldsymbol{\alpha}}\|}, \quad (2)$$

where $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}}$ are the ground truth and the recovered model parameters, respectively. Second, we define the RMSE as the average point-wise distance between the ground truth and the matched shape (or texture). Therefore we measure that the N_{iter}, T_{mat}, C_{shp}, C_{tex}, $RMSE_{shp}$, and $RMSE_{tex}$ are the average number of iterations, average computation time, normalized correlation of the shape parameters, normalized correlation of the texture parameters, average RMSE of the shape error, and average RMSE of the texture error, respectively. We used the BJUT-3D Face Database [1]. All experiments have been conducted on the desktop PC that consists of a Pentium IV 3Ghz, and C++ development tools.

4.1 Correlation of the Model Parameters among Different Layers

First, we evaluated the correlation of the shape and the texture parameters among two different levels. Since we take the model parameters obtained from the fitting result at the lower layer as the initial model parameters for the ICIA fitting at the upper layer, there should be a strong correlation between the model parameters of two adjacent levels. The correlation of the shape or texture parameters is evaluated by fitting the 3D face model to the face image at each level, and computation of the normalized correlation of the shape or texture parameters between two levels.

In the shape models, the correlations of the shape parameters among two different levels are 1 in all cases. It means that the correlations between the matched shape parameters of adjacent levels are a strong correlation.

Table 1 shows the normalized correlation of the fitted texture parameters between two levels, where they are different from each other. The normalized correlation between level1 and level2 is the highest one (0.9475), the normalized correlation between level2 and level3 is the next highest one (0.8476), and the the normalized correlation between level1 and level3 is the smallest but is still strong (0.7932). This also indicates that we can also use the fitted texture parameters at the lower layer as the initial texture parameters of the next upper layer sufficiently.

Table 1. The correlation of the texture parameters among different levels.

$C(\beta^*_{level_i}, \beta^*_{level_j})$	$\beta^*_{level_1}$	$\beta^*_{level_2}$	$\beta^*_{level_3}$
$\beta^*_{level_1}$	1	0.9475	0.7932
$\beta^*_{level_2}$	0.9475	1	0.8476
$\beta^*_{level_3}$	0.7932	0.8476	1

From these results, we can use the matched shape and texture parameters at the lower layer as the initial parameters of the next upper layer sufficiently.

4.2 Comparison of the Matching Performance with Different Types of Matching Methods

Second, we compared the matching performance with respective to two different types of matching method: the conventional ICIA matching method (CICIA) and the proposed hierarchical ICIA matching method (HICIA) when the GIP is used to generate the image pyramid and the TSPU is used for the parameter updates. Table 2 summarizes the matching performances of two matching methods. This table indicates that (1) the proposed HICIA matching method is faster than the conventional ICIA matching method by a speed up to 3, (2) C of shape and texture parameters of using HICIA are much higher than those of using CICIA, and (3) the RMSEs of the shape and texture of using HICIA are reduced by 3-4 times than those of using CICIA.

Table 2. Comparison of matching performance between CICIA and HICIA.

	N_{iter}	T_{mat}	C_{shp}	C_{tex}	R_{shp}	R_{tex}
CICIA	71.02	16.3	0.375	0.855	6.454	6.041
HICIA	57.52	5.794	0.820	0.993	2.167	1.605

Figure 1 shows a histogram of the shape errors between the ground truths and the matched shapes using CICIA and HICIA. It shows that HICIA has the smaller mean and the smaller standard deviation of the shape errors than CICIA.

Figure 2 compares the convergence rates of five different matching methods: TYPE 1 (CICIA), TYPE 2 (HICIA+SIP+SSPU), TYPE 3 (HICIA+SIP+TSPU), TYPE 4 (HICIA+GIP+SSPU), and TYPE 5 (HICIA+GIP+TSPU), which SIP and SSPU are the subsampling image pyramid and the single-stage model parameter updates respectively. In this experiment, the input face image is successfully converged when the shape error of the matching face image is smaller than

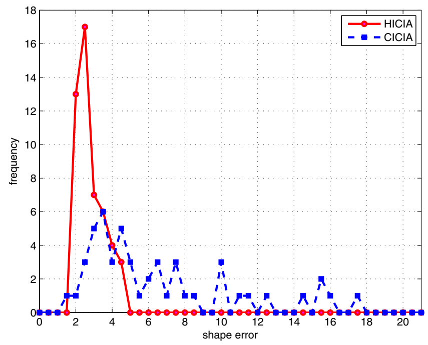

Figure 1. A histogram of the shape errors using the CICIA and the HICIA.

a given threshold value and the convergence rate is defined by the ratio of the number of successfully converged face images over the total number of face images. This figure illustrates that the convergence rate of the proposed HICIA method outperforms those of other matching methods.

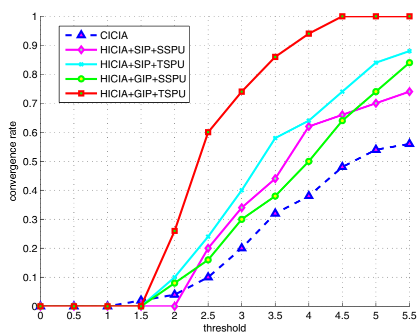

Figure 2. Convergence rates of the different matching methods.

5 Conclusion

We proposed a hierarchical ICIA matching method. The proposed HICIA is efficient because it generates the multi-resolution 3D face models and constructs the Gaussian image pyramid. Further the matching is conducted hierarchically from the lower layer to the higher layer. In addition, we also proposed the two-stage model parameter update method that updates the motion and texture parameters only at the first layer matching, and updated all parameters at the succeeding upper layer matching. We performed several experiments that validate the efficiency and accuracy of the proposed HICIA matching method. From the experiment results, (1) it completed the matching within about 57 iterations ($\approx$ 5 seconds), (2) its speed-up ratio is about 3, (3) the performance of the proposed matching method outperformed that of the existing matching method, (4) TSPU showed better matching performance than SSPU in terms of the matching time and matching error.

Acknowledgements

This work was supported by grant No. R01-2007-000-11683-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

References

[1] The BJUT-3D Large-Scale Chinese Face Database. Technical report, The Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory in Beijing University of Technology, 2005.

[2] J. Ahlberg. Using the active appearance algorithm for face and facial feature tracking. In *ICCV'01 Workshop on Recognition, Analysis and Tracking of Faces and Gestures in RealTime Systems*, pages 68–72, 2001.

[3] V. Blanz and T. Vetter. Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, 1999.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active Shape Models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] S. Gokturk, J. Bouguet, C. Tomasi, and B. Girod. Model-based face tracking for view-independent facial expression recognition. In *FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 272–278, 2002.

[6] R. Gross, I. Matthews, and S. Baker. Lucas-Kanade 20 Years on : A Unifying Framework: Part 3. Technical report, Carnegie Mellon University Robotics Institute, 2003.

[7] Y. Li, T. Gong, and H. Liddell. Modelling faces dynamically across views and over time. In *Proceedings of IEEE International Conference on Computer Vision*, 2001.

[8] A. W. M. Kass and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1987.

[9] I. Matthews and S. Baker. Active Apperance Models Revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[10] S. Romdhani and T. Vetter. Efficient, Robust and Accurate Fitting of a 3D Morphable Model. In *IEEE International conference on Computer Vision 2003*, pages 59–66, 2003.