Automatic Non-rigid Registration of 3D Dynamic Data for Facial Expression Synthesis and Transfer

Sen Wang, Xianfeng David Gu, Hong Qin
Stony Brook University (SUNY), Stony Brook, NY 11794
{swang, gu, qin}@cs.sunysb.edu

Abstract

Automatic non-rigid registration of 3D time-varying data is fundamental in many vision and graphics applications such as facial expression analysis, synthesis, and recognition. Despite many research advances in recent years, it still remains to be technically challenging, especially for 3D dynamic, densely-sampled facial data with a large number of degrees of freedom (necessarily used to represent rich and subtle facial expressions). In this paper, we present a new method for automatic non-rigid registration of 3D dynamic facial data using least-squares conformal maps, and based on this registration method, we also develop a new framework of facial expression synthesis and transfer. Nowadays more and more 3D dynamic, densely-sampled data become prevalent with the advancement of novel 3D scanning techniques. To analyze and utilize such huge 3D data, an efficient non-rigid registration algorithm is needed to establish one-to-one inter-frame correspondences. Towards this goal, a non-rigid registration algorithm of 3D dynamic facial data is developed by using least-squares conformal maps with additional feature correspondences detected by employing active appearance models (AAM). The proposed method with additional, interior feature constraints guarantees that the non-rigid data will be accurately registered. The least-squares conformal maps between two 3D surfaces are globally optimized with the least angle distortion and the resulting 2D maps are stable and one-to-one. Furthermore, by using this non-rigid registration method, we develop a new system of facial expression synthesis and transfer. Finally, we perform a series of experiments to evaluate our non-rigid registration method and demonstrate its efficacy and efficiency in the applications of facial expression synthesis and transfer.

1. Introduction and Previous Work

Automatic non-rigid registration of 3D time-varying data is a fundamental and enabling technique in 3D vision and graphics which has widespread applications. As 3D scanning technologies continue to improve, 3D dynamic densely-sampled data is becoming more and more prevalent for analysis, synthesis, and recognition purposes. To study and analyze such huge data, an efficient non-rigid registration algorithm is necessary to establish dense one-to-one inter-frame correspondences automatically. However, automatic 3D non-rigid registration still remains a challenging task, especially for dynamic densely-sampled facial expression data with many degrees of freedom. There has been much research on registration of 3D facial data in recent decades. Existing approaches to solving this problem typically involve three key techniques: one is to select feature correspondences manually or use markers attached to human faces [16, 21, 10]. The second one is to establish inter-frame correspondences hierarchically using multi-resolution facial data [28, 15]. The third kind of techniques computes correspondences using a low-resolution 3D deformable model [1, 17, 22]. However, most of these existing 3D non-rigid registration methods rely on recovering low dimensional parameters of face model or register 3D faces with local optimization that may not establish accurate one-to-one inter-frame correspondences successfully. In this paper, an automatic non-rigid registration algorithm of 3D dynamic densely-sampled facial data is developed using least-squares conformal maps with additional interior feature correspondences detected by active appearance model (AAM) [3, 8, 7]. The least-squares conformal maps between two 3D surfaces are globally optimized with less angle distortion and the resulting 2D map is stable, one-to-one, insensitive to resolution changes and robust in the existence of noise. Through the way of mapping 3D surfaces to a 2D common domain, it simplifies the original 3D surface-registration problem to a 2D registration problem. Thus, more accurate and efficient non-rigid registration algorithms could be achieved by using least-squares conformal maps. In sharp contrast to previous work on 3D non-rigid registration, especially the methods using attached markers, which unavoidably require much laborious human intervention and also are more invasive to human subjects, our new method can register non-rigid 3D dynamic data automatically and

efficiently with minimum manual work.

Conformal maps have already been employed in many vision and graphics applications most recently. A surface matching method based on harmonic maps was proposed in [30]. Sharon et al. [20] use conformal maps to analyze similarities of 2D shapes. Moreover, conformal maps are used for 3D face and brain surface matching in [9, 26]. Least-squares conformal maps are introduced by Levy et al. [14] for texture atlas generation and used by Wang et al. [24, 25] to conduct 3D surface matching with feature detection using the technique of spin-image. Because spin-image can only detect features on surfaces with rigid transformation, their method can not guarantee to successfully match surfaces with non-rigid deformation. For non-rigid 3D surface registration, Wang et al. [27] use a modified harmonic map to track 3D high resolution facial motion data. In order to calculate these harmonic maps, the surface boundary must be identified and a boundary mapping from 3D surfaces to the 2D domain must be properly created, which can be a difficult task especially when parts of the surface are occluded. In contrast, conformal maps and least-squares conformal maps do not necessarily require boundary information to be aligned and so give rise to a natural choice to combat this difficulty. Moreover, least-squares conformal maps enable users to enforce more interior feature constraints which will guarantee to achieve more accurate registration results in an automatic way.

Realistic facial animation and expression analysis remains a central challenge in vision and graphics. Earlier approaches explicitly model the facial anatomy, deriving facial animations from the physical behaviors of the bone, joint, and muscle structures [13, 29]. Others focus only on the surface of the face, using smooth surface deformation mechanisms to create facial expressions [10, 16, 28]. These approaches make use of existing data for animating a new model. Previous works also use techniques for tracking head motions and facial expression in video [4, 17] and copy deformations from one subject onto the geometry of other faces [1]. Expression cloning [16, 21, 18] improves upon this deformation transfer process with both 3D source and target face data. Recently, facial animation and expression analysis using 3D motion capture becomes available with the advancement of new 3D scan techniques [31, 28]. However, these 3D motion data is not registered in the space-time domain. For this purpose, a number of registration methods have been proposed for 3D dynamic facial data. Zhang et al. [31] propose a new tracking method based on optic flow estimation which can be sensitive to noise. Wang et al. [28] use a hierarchical method to track 3D motion facial data with expression transfer at the cost of making the estimation of model parameters more difficult. Moreover, their method requires a lot of manual work by dividing the face model into several deformable regions.

Facial expression undergos complicated global and local nonlinear deformation between frames and is represented by a high dimensional vector (a collection of 3D vertices). It is impossible to analyze and synthesize facial expression in high dimensional space. In this paper, we describe a dynamic facial expression synthesis system using isomap[23] which can embed facial expression manifolds in high dimension into low dimensional space. Finally, we present a facial expression transfer framework based on our non-rigid registration method using least-squares conformal maps and our approaches lead to more accurate results with minimum human intervention.

The rest of the paper is organized as follows: The theoretical background of least-squares conformal map is introduced in Section 2. A non-rigid registration method of 3D dynamic facial data using least-squares conformal maps is documented in Section 3. A system of facial expression synthesis and transfer is presented in Section 4. Experimental results are discussed in Section 5, and we conclude our paper and point out future work in Section 6.

2. Theoretical Background of Least-Squares Conformal Map

From Riemann theory, all 3D surfaces are Riemann surfaces and have unique conformal structures. Conformal structure is more flexible than Riemannian metric structure and more rigid than topological structure. By applying conformal structure to surface analysis, any 3D surface can be mapped onto a 2D domain through a global optimization with angle-preserving and the mappings are easy to control by specifying a few number of landmarks. Because of this important property of conformal geometric maps, it can simplify all 3D problems to 2D ones. Moreover, it can be proven that there exists a conformal map from any surface with a disk topology to a 2D planar domain, which is a diffeomorphism, namely, *one-to-one*, and *onto*. Conformal parameterization depends on the geometry itself, not the triangulation of the surfaces. In this section, we introduce the notion of conformal map and least-squares conformal map with relationship between them.

2.1. Conformal Map

Conformal map has deep relations with complex geometry. In planar case, it becomes complex analysis [6, 14]. Suppose we map a planar region S to the plane, the map is denoted as $\mathcal{U} : (x, y) \rightarrow (u, v)$. For convenience, we use complex coordinates $z = x + iy, \mathcal{U} = u + iv$. Then $\mathcal{U}(z)$ is a complex function. According to complex function theorem, a conformal map is equivalent to an analytic function, which satisfies the Riemann-Cauchy equation

$$\frac{\partial \mathcal{U}}{\partial \bar{z}} = 0, \tag{1}$$

where the complex differential operator $\frac{\partial}{\partial \bar{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$.

Suppose both the domain and the target are the unit disk, then all conformal maps form a three dimensional group, each of them is called a Möbius transformation. A Möbius transformation has the form: $\mathcal{U} = \frac{z - z_0}{1 - \bar{z}_0 z} e^{i\theta}$, where z_0 is the pre-image of the center of the disk, θ is an angle in $[0, 2\pi]$. For general surfaces which are homeomorphic to a disc, Riemann's theorem states that the conformal maps satisfying Eq. (1) exists and if the target planar domain is fixed, all such mappings form a 3-dimensional group. Furthermore, by specifying one interior point and one boundary point, the conformal map is uniquely determined [9].

However, in practice, since our goal is to use the resulting conformal parameterization for registration, we need to introduce additional feature constraints, indicating that the corresponding features on two 3D surfaces should be mapped onto the same locations in the 2D domain. However, with these additional constraints, it is not always possible to satisfy the conformality condition. For this reason, we seek to minimize the violation of Riemann's condition in the least-squares sense.

2.2. Least-Squares Conformal Map

The Least-Squares Conformal Map (LSCM) parameterization algorithm generates a discrete approximation of a conformal map by adding more constraints[14, 24, 25]. Given a discrete 3D surface mesh S and a piecewise linear map $\mathcal{U} : S \rightarrow \mathbb{R}^2$, as described in Section 2.1, $\mathcal{U}$ is conformal on S if and only if the Cauchy-Riemann equation (Eq. (1)) holds true on the entire surface S. However, by adding additional constraints, this conformal condition cannot be strictly satisfied on the entire triangulated surface S, so the conformal map is constructed in the least-squares sense. Firstly, we measure the conformality of this planar linear map $\mathcal{U}$ on each triangle d using $|\frac{\partial \mathcal{U}}{\partial \bar{z}}|^2 A(d)$, where $A(d)$ is the area of the face d. The conformality energy of the entire map $\mathcal{U} : S \rightarrow \mathbb{R}^2$ is just the total energy on all triangles of the surface and the LSCM can be achieved by

$$MinC(S) = Min \sum_{d \in S} |\frac{\partial \mathcal{U}}{\partial \bar{z}}|^2 A(d). \qquad (2)$$

Let $\alpha_j = u_j + iv_j$ and $\beta_j = x_j + iy_j$, so $\alpha = \mathcal{U}(\beta)$. Because $\mathcal{U}$ is piecewise linear, therefore, $\mathcal{U}$ can be represented as a matrix. Then, we rearrange the vector α such that $\alpha = (\alpha_f, \alpha_p)$ where α_f consists of $n - p$ free coordinates and α_p consists of p constraint point coordinates. Therefore, Eq. (2) can be rewritten as

$$C(S) = \|M_f \alpha_f + M_p \alpha_p\|^2, \qquad (3)$$

where $M = (M_f, M_p)$, a sparse $m \times n$ complex matrix (m is the number of triangles and n is the number of vertices). The least-squares minimization problem in Eq. (3) can be efficiently solved using the Conjugate Gradient

Method. Thus we can map a 3D surface to a 2D domain with multiple correspondences as constraints by using the LSCM technique.

As described above, LSCMs are generated by minimizing the violation of Riemann's condition in the least-squares sense. This optimization-based parameterization method has the same properties as conformal maps, e.g., existence and uniqueness and also can map a 3D shape to a 2D domain in a continuous manner with minimized local angle distortion. Moreover, LSCMs are independent of mesh resolution and can handle missing boundaries and also allow multiple constraints to be enforced. Finally, the least-squares minimization problem in calculating LSCMs has the advantage of being linear. With such good properties, we expect them to be very valuable in 3D surface registration.

3. Non-rigid Registration Algorithm for 3D Dynamic Facial Data

We now introduce an automatic non-rigid registration algorithm by using least-squares conformal maps which can map 3D surfaces to a 2D common domain with global optimization. Therefore, they can simplify the original 3D surface-registration problem to a 2D registration problem. In particular, our registration algorithm includes two steps: First, interior feature correspondences are detected by using Active Appearance Model (AAM); After that, by generating and registering the 2D least-squares conformal maps of 3D faces in two frames, we compute their dense one-to-one correspondences to register these two frames.

3.1. Feature Tracking

There are many features in the human face such as corners of eyes, nose and mouth. Detecting and tracking these features accurately and efficiently in 3D dynamic facial data still presents difficulties. Active Appearance Model (AAM)[3, 8, 7] is successfully used to track facial features in video sequences. AAM is a face detection technique that combines shape and texture information into one PCA space. The model iteratively searches a new image by using the texture residual to update the model parameters. To use AAM to detect features in 3D dynamic facial data, firstly, we use a projection matrix P to project the 3D faces onto a 2D image plane. Then we use AAM to detect the features in each 2D video frame. After that, with the known projection and depth information of these 3D data, we can project the features detected by AAM back to 3D face surfaces. Finally, we can automatically get the initial inter-frame feature correspondences in these 3D dynamic data. In experiments, we select 200 frames in training data containing different facial expressions to build the AAM and the facial feature template contains 50 vertices, as shown in Fig. 1.
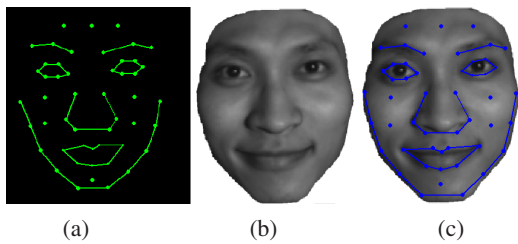
Figure 1. AAM feature detection. (a) The feature template of AAM. (b) A 3D face projected onto an image plane. (c) The detected features on the face.
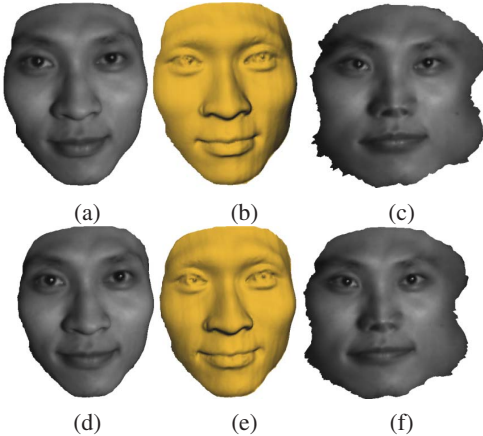

Figure 2. Registration using least-squares conformal maps (LSCMs). (a) and (d) are two original inter-frame 3D face surfaces with texture information. (b) and (e) are these faces without texture. (c) and (f) are their registered LSCMs.

3.2. Dynamic Registration Using Least-Squares Conformal Maps

After detecting the initial corresponding features in two frames S_i and S_{i+1}, we can compute their least-squares conformal maps (LSCMs) using the method described in Section 2.2. As the LSCMs are driven by representative motion features between the two frames, they capture the inter-frame non-rigid deformation. Furthermore, because this mapping is *one-to-one and onto*, by registering their 2D LSCMs, we can recover the inter-frame registration on these 3D face surfaces.

As an example, Fig. 2(c,f) show the LSCMs of the inter-frame 3D faces in Fig. 2(a,d). Fig. 2(a,d) are the original faces with texture information and Fig. 2(c,f) are their registered 2D LSCMs. The similarity of these two LSCMs in Fig. 2(c,f) shows that we can successfully register two inter-frame non-rigid 3D faces by just registering their 2D LSCMs.

4. A Framework of Facial Expression Synthesis and Transfer

We now present the new framework of dynamic facial expression synthesis and transfer based on our non-rigid registration method.
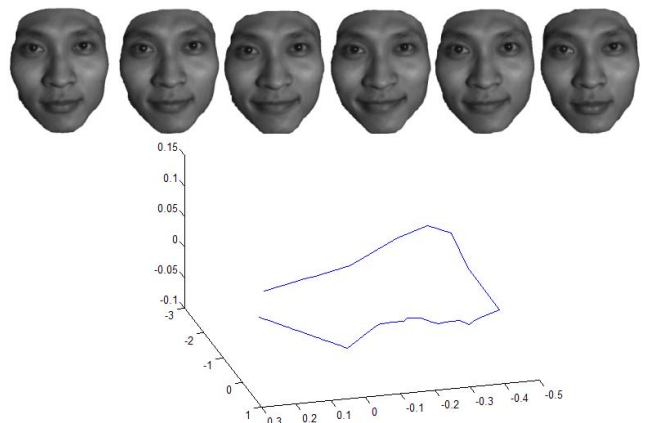

Figure 3. Facial expression manifold. The curve is Isomap for 3D registered facial expression sequence(some frames are shown in first row).

4.1. Dynamic Facial Expression Synthesis

Expression synthesis generates new facial animations using existing expression data. Our expression synthesis framework includes two steps: The first step analyzes existing expression data by embedding them into a low-dimensional manifold using Isomap [23] after registering these data using our 3D non-rigid registration method described in section 3. The second step synthesizes new expressions by selecting parameters of these expression data analyzed in the first step.

4.1.1 Facial Expression Manifold Embedding

Facial expression undergos complicated global and local nonlinear deformation between frames. In order to analyze expression data easily and efficiently, we need to embed facial expression manifolds non-linearly into a low dimensional space. We adapt Isomap framework [23] to achieve a low dimensional manifold embedding for individual facial expressions that provides a good representation of facial motion. Isomap finds the best embedding manifold with nonlinear dimensionality reduction by preserving the proportion of distance in the embedding space and the original facial motion space. Fig. 3 shows the embedding of smile motion to a 3D space. It is an elliptical one dimensional manifold in 3-dimensional space. In embedding space, the expression manifolds are elliptical curves with distortions according to face geometry, expression types. To analyze these expression manifolds, we need to align these one dimensional manifolds in embedding space. For each manifold, correspondences are initially established using the points with high curvatures. Then, multiple manifolds are aligned using an approach similar to [2]. Thus, we can align the original expression sequences in temporal space by aligning expression manifolds in the embedding space.

4.1.2 Expression Synthesis

After we align N expression styles $s_1, s_2, ..., s_n$ of the same person using the method described above, we then generate a new style vector s_{new} by linear interpolation of these N styles using control parameters $w_1, w_2, ..., w_n$ as follows:

$$s_{new} = w_1 s_1 + w_2 s_2 + ... + w_n s_n \qquad (4)$$

where $\sum_{i=1}^{N} w_i = 1$. For example, if we want to generate new expression as style with 50% of the first style and 30% of the second style and 20% of the third style, then we generate new style as $s_{new} = 0.5 s_1 + 0.3 s_2 + 0.2 s_3$.

4.2. Facial Expression Transfer

Expression transfer directly maps expressions of the source model to the target model. In particular, our expression transfer framework includes two steps: The first step determines temporal correspondences between every two adjacent frames of the source model and spatial correspondences between the source and target models; The second step transfers the adjusted motion vectors from source model vertices to target model vertices.

4.2.1 Dense Surface Correspondences

Source models at each frame do not have temporal interframe correspondences. In addition, source model and target model do not have spatial correspondences as they may have different structures. However, we can establish both temporal and spatial correspondences by using parameterization methods [5, 6, 11, 14] to map 3D source and target models to a 2D domain. Therefore, we can compute 3D dense surface correspondences by just detecting correspondences in their 2D maps.

Temporal Correspondences: In our experiments, we use fine facial motion data which are captured by a structured lighting method [32] with 30 frame per second. A 3D face in each frame has approximately 70K points with both shape and texture information. To utilize this 3D dynamic data, we use our 3D non-rigid registration method described in Section 3 to obtain the one-to-one inter-frame correspondences, as shown in Fig. 2.

Spatial Correspondences: For expression transfer, it is crucial to find spatial correspondences between the source and target models. Harmonic mapping is a popular approach for recovering dense surface correspondences [5, 12]. However, difficulties arise when specific points need to be matched exactly between models. Our approach to finding spatial correspondences starts with initial corresponding feature points which the user specifies [12] between the source and target models. After that, we simplify the source and target models and map them to a 2D plane by minimizing the harmonic energy [5, 9, 30] with user-specified corresponding feature points as interior constraints. By detecting and interpolating the one-to-one correspondences in

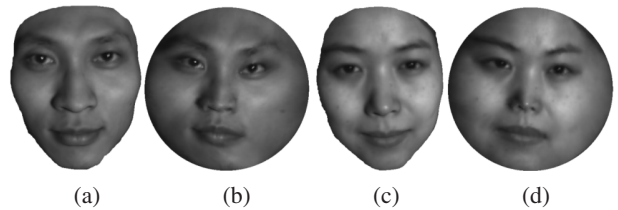

(a)   (b)   (c)   (d)

Figure 4. Spatial feature correspondence detection using harmonic maps. (a) and (c) are source and target faces. (b) and (d) are their harmonic maps computed by our method. After detecting the one-to-one correspondences in their 2D harmonic maps, we can obtain the spatial feature correspondences between 3D source and target faces.
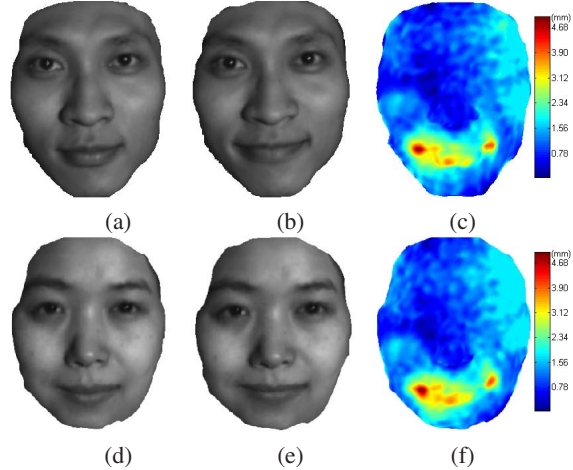

(a)   (b)   (c)

(d)   (e)   (f)

Figure 5. An example of motion vector transfer. (a) and (b) are source faces with different expressions. (c) is the color-coded magnitude of motion vectors in the source model. (d) is the target face model. (e) is transferred expression on the target face. (f) is the color-coded magnitude of motion vectors to be transferred to the target face model (d).

the 2D harmonic maps, we can obtain the spatial correspondences between the source and target models, as shown in Fig. 4.

4.2.2 Expression Transfer with Motion Vectors

A transferred expression animation displaces each target vertex to match the motion of a corresponding surface point in source model. Since facial geometry and aspect ratios are different between the scans of source models and the target face, source displacement vectors can not be simply transferred without adjusting the direction and magnitude of each motion vector. In our experiments, we adjust both the scale and orientation of motion vectors before transferring the source motion to target model by using the method described in [16]. An example of motion vector transfer is shown in Fig. 5.

5. Experimental Results

The performances of our approaches on non-rigid registration of 3D time-varying data and facial expression syn-

thesis and transfer are evaluated in a number of experiments. First, we analyze the accuracy of our 3D non-rigid registration method and compare results with two previous methods. Second, we evaluate the performance of facial expression synthesis and transfer based on our non-rigid registration method.

5.1. Evaluation of 3D Non-rigid Registration

We apply our non-rigid registration method on 3D dynamic facial data and compare results with the tracking method based on modified harmonic maps [27] and Iterative Closest Point (ICP) method [19] which have been widely used for 3D registration. In order to evaluate their accuracy, we compute the registration error by approximately using the difference in the intensity values of vertices of registered 3D face surfaces between two frames as:

$$RegistrationError = \frac{\sum_{i=1}^{N} \|t_j^i - t_{j+1}^i\|}{\sum_{i=1}^{N} t_j^i}, \qquad (5)$$

where t_j^i is intensity value of the ith vertex of 3D face surface in the jth frame and N is the number of registered vertices. If the registration is perfect, the only difference in the intensity values of vertices of registered two 3D faces will result from the change of shadowing and shading effects due to geometric deformation.

We present the comparison of these three techniques in Fig. 6 by plotting the registration errors according to different frames. From the results, we can see that our method performs considerably better than the other two methods. The ICP method can not achieve good results in 3D non-rigid shape registration. The modified harmonic map method uses optical flow to track very few feature points which are very sensitive to noise. Moreover, their method will have larger registration errors in the 3D face data with varying boundary, because of the limitation of harmonic maps.

5.2. Evaluation of Facial Expression Synthesis and Transfer

Firstly, We apply our facial expression synthesis framework on 3D dynamic facial data to synthesize new facial expressions. Actors perform four different type of expressions: smile, surprise, sadness and anger. The expressions were captured using our structured lighting ranger scanner. We then registered and analyzed the captured ranger data using our facial expression synthesis framework described in section 4.1. Fig. 7 shows the generation of two expressions: smile and surprise and the synthesis of a new in-between expression by changing the weight of these two original input expressions. With our method we can generate a convincing combination of two different expressions without loss of details. The generated in-between expressions are shown in the second and third rows.
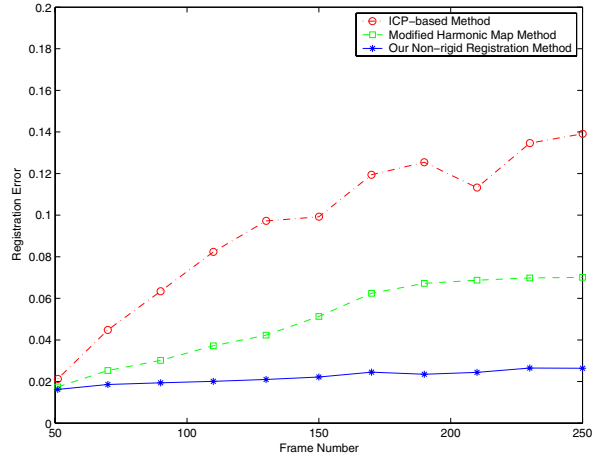

Figure 6. Comparison of the three registration methods.

Figure 7. Synthesis of new facial expression by weighting two different expression types: smile in the first row and surprise in fourth row. Second row: $70\%smile + 30\%surprise$. Third row: $30\%smile + 70\%surprise$.

Next, we apply our facial expression transfer framework on facial data with different expressions and transfer these expression styles and details to target face models. We perform two group experiments to evaluate the accuracy and robustness of our facial expression transfer method both qualitatively and quantitatively. Our first group experiments are intended to qualitatively show the effectiveness of our expression transfer approach. Fig. 8 shows the expression transfer results with various exaggerated expressions and Fig. 9 shows the results with different kinds of expressions which are neutral, smile, surprise, sadness and anger. We also perform expression transfer from the source model to a topologically different target face model caused by miss-
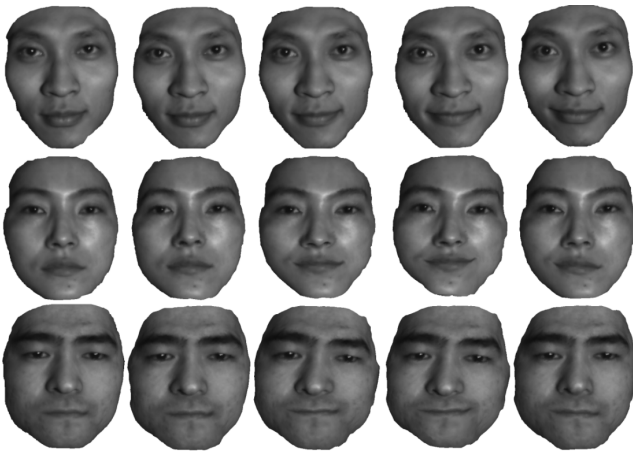
Figure 8. Exaggerated expression transfer. Source face model with exaggerated expressions are shown in the first row. Transferred expressions on two target faces are showed in the second and third row, respectively. The target faces have different shapes and textures but the expressions are proportionally scaled to fit each model well.

Figure 9. Expression transfer. Source face model with different expressions are shown in the first row. Transferred expressions on the target face are shown in the second row. From left to right, emotional expressions are neutral, happy, surprised, sad and angry, respectively.

ing data in eye regions during data acquisition under different resolutions and the results are shown in Fig. 10. As shown in these results, the expressions of the source model are reproduced in the target model with convincingly better effects.

The second group experiments are intended to quantitatively measure the effectiveness of our expression transfer approach. In the third experiment, we use two different 3D scans of the male subject in Fig. 9 as source and target models, respectively, that is, transferring expressions from a person to himself. In the last experiment, we transfer expressions of the male subject to the female subject in Fig. 9 and then transfer intermediate results back to another 3D scan of the male subject. By using Eq. (5), the average errors of intensities are measured between the original and final face models in all frames as shown in Table 1. Fig. 11 exhibits some of these expression transfer results in different frames. From the results, we can see that in each frame, the final

Figure 10. Expression transfer from a male subject to a topologically different face model under different resolutions. Source face model with different expressions are shown in the first row. Transferred expressions on the target face which has different topology due to missing data (missing in eye region during data acquisition) are shown in the second row. Expression transfer results of the target face with only 1/4 of the original resolution are shown in the third row.

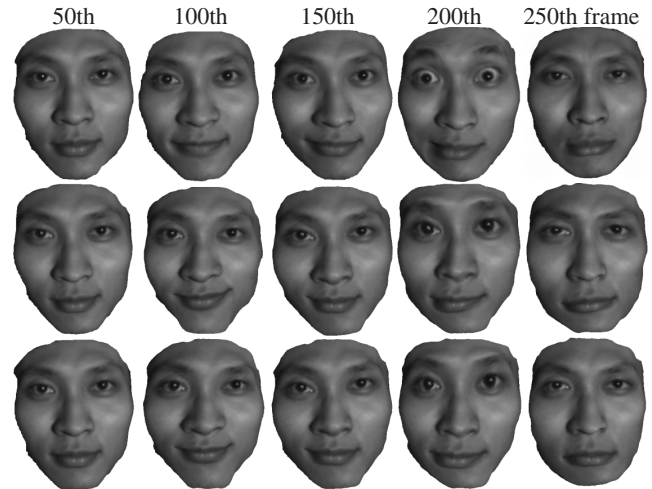50th   100th   150th   200th   250th frame

Figure 11. Expression Transfer results (Man $\Rightarrow$ Man and Man $\Rightarrow$ Woman $\Rightarrow$ Man). Source face models in different frames are shown in the first row. Expression transfer results (Man $\Rightarrow$ Man) are shown in the second row. Expression transfer results (Man $\Rightarrow$ Woman $\Rightarrow$ Man) are shown in the third row.

faces after expression transfer are very similar to the original source face data and the only difference results from the change of the shadowing and shading effects due to face geometry deformation. The overall processing time including 3D non-rigid registration and expression transfer is approximately 1 minute per frame on a Pentium4 2.4 GHz PC. From all of these results, comparing with the previous research on expression transfer which typically require many manual labors, our method can transfer expression from one person to another efficiently and automatically.

Table 1. Average errors of expression transfer.

	Man ⇒ Man	Man ⇒ Woman ⇒ Man
Average RegistrationError	2.312%	2.379%

6. Conclusion and Future Work

We have developed a novel method for non-rigid registration using least-squares conformal maps to automatically compute dense one-to-one inter-frame correspondences for 3D time-varying facial data. Moreover, based on this registration method, we have also implemented a new visual modeling framework of dynamic facial expression synthesis and transfer. Our experimental results demonstrate that our novel facial modeling framework leads to better registration for 3D dynamic facial data and subsequent applications such as facial expression analysis, synthesis and transfer.

Regarding on-going and near-future work, we shall further validate our methods and our framework on a much wider range of 3D dynamic facial data through more extensive testings. We also plan to use our facial expression synthesis and transfer framework for other vision applications such as facial expression classification and recognition.

References

[1] J. Chai, J. Xiao, and J. Hodgins. Vision-based control of 3d facial animation. In *Symposium on Computer Animation*, pages 193–206, 2003. 1, 2

[2] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Comput. Vis. Image Underst.*, 89(2-3):114–141, 2003. 4

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV98*, pages 484–498, 1998. 1, 3

[4] D. DeCarlo and D. Metaxas. The integration of optical flow and deformable models with applications to human face shape and motion estimation. In *CVPR96*, page 231, 1996. 2

[5] M. Eck, T. DeRose, T. Duchamp, H. Hoppe, M. Lounsbery, and W. Stuetzle. Multiresolution analysis of arbitrary meshes. In *SIGGRAPH95*, pages 173–182, 1995. 5

[6] M. S. Floater and K. Hormann. Surface parameterization: a tutorial and survey. In *Advances in Multiresolution for Geometric Modelling*, pages 157–186. 2004. 2, 5

[7] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image and Vision Computing*, 23(11):1080–1093, November 2005. 1, 3

[8] R. Gross, I. Matthews, and S. Baker. Active appearance models with occlusion. *Image and Vision Computing*, 24(6):593–604, 2006. 1, 3

[9] X. Gu, Y. Wang, T. F. Chan, P. M. Thompson, and S. Yaun. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transaction on Medical Imaging*, 23(7):949–958, 2004. 2, 3, 5

[10] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making faces. In *SIGGRAPH98*, pages 55–66, 1998. 1, 2

[11] S. Haker, S. Angenent, A. Tannenbaum, R. Kikinis, G. Sapiro, and M. Halle. Conformal surface parameterization for texture mapping. *IEEE Transactions on Visualization and Computer Graphics*, 6:181–189, 2000. 5

[12] A. W. F. Lee, D. Dobkin, W. Sweldens, and P. Schroder. Multiresolution mesh morphing. In *SIGGRAPH99*, pages 343–350, 1999. 5

[13] Y. Lee, D. Terzopoulos, and K. Walters. Realistic modeling for facial animation. In *SIGGRAPH95*, pages 55–62, 1995. 2

[14] B. Levy, S. Petitjean, N. Ray, and J. Maillot. Least squares conformal maps for automatic texture atlas generation. In *SIGGRAPH02*, pages 362–371, 2002. 2, 3, 5

[15] K. Na and M. Jung. Hierarchical retargetting of fine facial motions. In *EUROGRAPHICS04*, pages 687–695, 2004. 1

[16] J. Y. Noh and U. Neumann. Expression cloning. In *SIGGRAPH01*, pages 277–288, 2001. 1, 2, 5

[17] F. Pighin, R. Szeliski, and D. Salesin. Resynthesizing facial animation through 3d model-based tracking. In *ICCV99*, pages 143–150, 1999. 1, 2

[18] H. Pyun, Y. Kim, W. Chae, H. W. Kang, and S. Y. Shin. An example-based approach for facial expression cloning. In *SCA03*, pages 167–176, 2003. 2

[19] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Realtime 3d model acquisition. In *SIGGRAPH02*, pages 438–446, 2002. 6

[20] E. Sharon and D. Mumford. 2d shape analysis using conformal mapping. In *CVPR04*, pages II: 350–357, 2004. 2

[21] R. W. Sumner and J. Popovic. Deformation transfer for triangle meshes. In *SIGGRAPH04*, pages 399–405, 2004. 1, 2

[22] H. Tao and T. Huang. Explanation-based facial motion tracking using a piecewise bezier volume deformation model. In *CVPR99*, pages I: 611–617, 1999. 1

[23] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, Dec 2000. 2, 4

[24] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras. 3d surface matching and recognition using conformal geometry. In *CVPR06*, pages II: 2453–2460, 2006. 2, 3

[25] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras. Conformal geometry and its applications on 3d shape matching, recognition, and stitching. *IEEE Transection on Pattern Analysis and Machine Intelligence*, 29(7):1209–1220, 2007. 2, 3

[26] Y. Wang, M. C. Chiang, and P. M. Thompson. Mutual information-based 3d surface matching with applications to face recognition and brain mapping. *ICCV05*, 1:527–534, 2005. 2

[27] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras, and P. Huang. High resolution tracking of non-rigid 3d motion of densely sampled data using harmonic maps. In *ICCV05*, pages I: 388–395, 2005. 2, 6

[28] Y. Wang, X. Huang, C. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dyanmic 3-d facial expressions. In *Computer Graphics Forum*, pages III: 677–686, 2004. 1, 2

[29] K. Waters. A muscle model for animation three-dimensional facial expression. In *SIGGRAPH87*, pages 17–24, 1987. 2

[30] D. Zhang and M. Hebert. Harmonic maps and their applications in surface matching. In *CVPR99*, pages II: 524–530, 1999. 2, 5

[31] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23(3):548–558, 2004. 2

[32] S. Zhang and P. Huang. High resolution, real time 3d shape acquisition. In *CVPR04 Workshop on Real-time 3D Sensors and Their Use*, page 28, 2004. 5