

A Hierarchical Framework For High Resolution Facial Expression Tracking

Xiaolei Huang¹, Song Zhang², Yang Wang³, Dimitris Metaxas¹, Dimitris Samaras³

¹ Computer Science Department, Rutgers University – New Brunswick, NJ, USA
{xiaolei, dnm}@cs.rutgers.edu

² Mechanical Engineering Department, State University of New York at Stony Brook, NY, USA
szhang@oml.eng.sunysb.edu

³ Computer Science Department, State University of New York at Stony Brook, NY, USA
{yangwang, samaras}@cs.sunysb.edu

Abstract

We present a novel hierarchical framework for high resolution, nonrigid facial expression tracking. The high quality dense point clouds of facial geometry moving at video speeds are acquired using a phase-shifting based structured light ranging technique [19]. To use such data for temporal study of the subtle dynamics in expressions and for face recognition, an efficient nonrigid facial tracking algorithm is needed to establish intra-frame correspondences. In this paper, we propose such an algorithmic framework that uses a multi-resolution 3D deformable face model, and a hierarchical tracking scheme. This framework can not only track global facial motion that is caused by muscle action, but fit to subtler expression details that are generated by highly local skin deformations. Tracking of global deformations is performed efficiently on the coarse level of our face model with one thousand nodes, to recover the changes in a few intuitive parameters that control the motion of several deformable regions. In order to capture the complementary highly local deformations, we use a variational algorithm for non-rigid shape registration based on the integration of an implicit shape representation and the Free Form Deformations (FFD). Due to the strong implicit and explicit smoothness constraints imposed by the algorithm, the resulting registration/deformation field is smooth, continuous and gives dense one-to-one intra-frame correspondences. User-input sparse facial feature correspondences can also be incorporated as hard constraints in the optimization process, in order to guarantee high accuracy of the established correspondences. Extensive tracking experiments using the dynamic facial scan of five different subjects demonstrate the accuracy and efficiency of our proposed framework.

1. Introduction

Facial expression tracking is a fundamental problem in computer vision due to its important role in a variety of applications including facial expression recognition, classification, detection of emotional states, among others.

Most facial motion or expression tracking algorithms in the literature utilize image data from 2D video sequences [3, 4, 5, 10, 16], and focus on the accurate tracking of a low number of facial features such as points located around the brows, eyes, nose, mouth, etc. While the movements of these feature points in an expression can often be used effectively in classification, they are hardly sufficient in most recognition applications, since many distinct characteristics of a person's expression lie in the subtle details such as wrinkles and furrows. In video sequences, it can be very difficult to capture these details due to the loss of information in projection, lighting, shadow, and other conditions.

Recent technological advances in digital imaging, digital projection display, and personal computers, are making 3D shape acquisition in real time increasingly available. Such ranging techniques include structured light [7, 14], and spacetime stereo [18, 2]. These systems can capture 3D data quickly. Recently, Zhang and Huang developed a high-resolution dynamic 3D expression data acquisition system that can capture highly accurate geometry at speeds that exceed video frame rate [19]. Such high-quality data is very attractive in analysis of facial expressions. However, the dense data samples in these 3D face scans are not registered in object space and hence there is no guarantee of intra-frame correspondences, which would make tracking of facial features, temporal study of facial expression dynamics and other analysis difficult.

In the literature, few tracking algorithms have been proposed for 3D facial expression data. Existing methods pursuing 3D intra-frame correspondences largely fall into two categories: One depends on markers that are attached on the

face performing an expression [6] or on facial feature correspondences manually selected by users [17]; the other uses 3D shape registration algorithms such as [1, 20] to establish correspondences based on facial geometry. However, when used for tracking, previous methods in both categories lack a proper modelling of the motion style in an expression, which results from the combined effect of global facial motion that is caused by muscle action, and subtler expression details such as wrinkles that are generated by highly local skin deformations.

In this paper, we propose a novel hierarchical tracking framework for 3D dynamic expression data, which can both track global facial motion and fit to expression details, providing a tight coupling between the global and local deformations. The high quality moving face range scans we use are acquired using the system described by [19]. In order to track facial features and establish dense intra-frame correspondences, we use a multi-resolution deformable face model. On the coarse level we use a mesh with one thousand (1K) nodes that was first developed for robust face tracking in low quality 2D images [5] and extend it to deal with 3D range data. This method is fast, and the deformation parameters for each facial motion are few and intuitive. However it cannot capture the large amount of local deformations and so we use it for a coarse-level tracking. The local deformations and details in expressions are captured in a higher level fitting process. For each frame of the range scan, the resulting mesh from the coarse-level tracking is used to initialize a subdivided fine mesh with sixteen thousand (16K) nodes. This fine mesh is registered to the frame based on the 3D extension of a variational algorithm for non-rigid shape registration [8]. This algorithm integrates an implicit shape representation [12] and the cubic B-spline based Free Form Deformations (FFD) model [15, 13], and generates a registration/deformation field that is smooth, continuous and gives dense one-to-one correspondences.

Using our hierarchical framework, we did tracking experiments on dynamic facial scan of five different subjects, and conducted both qualitative and quantitative validation on the tracking accuracy. The results are very promising, showing the potential of our algorithm to serve as an efficient way to parameterize high resolution 3D dynamic expression data in order to make it easy to use while preserving the accuracy and visual quality that such data guarantees.

The remainder of the paper is organized as follows: In Section 2, we give an overview of the hierarchical tracking framework. Section 3 deals with tracking of the global deformations, and Section 4 describes the local registration algorithm between face scans. Experimental results are presented in Section 5, and we conclude in Section 6.

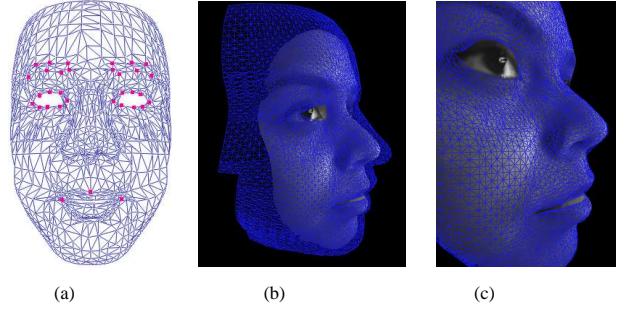


Figure 1. (a) The generic face model with manually selected feature points. (b) The face model and the face scan data are roughly aligned. (c) The result of the initial fitting to a 3D face scan data.

2. Overview: A Hierarchical Tracking Framework

To track the facial motion in an expression, we use a multi-resolution deformable face model. The face model has two resolutions: a coarse-level mesh with 1K nodes and a fine-level mesh with 16K nodes.

We use the 16K node mesh for the initial fitting between the face model and an actor’s face scan before performing an expression (i.e., the first frame). Figure 1 demonstrates this initial fitting process. First, the face model and the 3D scan data are roughly aligned by hand (Figure 1(b)). Then the 3D extension of a variational non-rigid shape registration algorithm [8] is used to register the face model with this range scan, achieving a complete surface match (Figure 1(c)). The algorithm is based on the integration of an implicit shape representation and cubic B-spline based Free Form Deformations (FFD). It represents shapes (surfaces) in an implicit form by embedding them in the space of distance functions of a Euclidean metric. A cubic B-spline based Free Form Deformation (FFD) model [15, 13] is then used to minimize a sum-of-squared-differences criterion between the embedding functions of a source and a target surface, and recover the FFD parameters that would map the source to the target. In this paper, in order to constrain the initial dense correspondences established by the registration algorithm, we define a small set of feature points on the face model (typically around 30, see Figure 1(a) as an example), then manually select their correspondences on the range data. These feature correspondences are incorporated as hard constraints during the optimization process of the registration algorithm (see section 4 for details). As a result, the initial correspondences established, especially between facial features such as tip of the nose and corners of the eyes, are very good.

After the initial fitting, a hierarchical scheme is adopted to track the intra-frame deformations in an expression. In

the coarse level, we use the 1K node face model and extend the deformable tracking system in [5] to track 3D dynamic range scans. In order to fit to expression details, for each frame of the range data, we use the coarse level tracking result to initialize the subdivided 16K node mesh in a higher level. Then this 16K node refined mesh is registered to the frame using the same variational non-rigid shape registration algorithm used for initial fitting. This hierarchical tracking/fitting protocol provides a tight coupling between global and local deformations, and results in efficient and very detailed fitting to the 3D face scan data (see Figure 2 for examples).

In the next two sections, we give details on our hierarchical 3D tracking approach, including the global tracking and local registration algorithms.

3. Global deformation

According to [11], we can express the position $\mathbf{p}(t)$ of a point on the 3D model as the sum of a reference model $\mathbf{s}(t)$ and a displacement $\mathbf{d}(t)$, i.e.

$$\mathbf{p} = \mathbf{s} + \mathbf{d} \quad (1)$$

The shape, position, and orientation of the reference model \mathbf{s} can also change. We define the reference shape as

$$\mathbf{s} = \mathbf{T}(\mathbf{q}; \mathbf{e}) \quad (2)$$

where \mathbf{s} is the result of a geometric primitive \mathbf{e} undergoing the *global deformation* \mathbf{T} . \mathbf{T} depends on a set of n control parameters $\mathbf{q} = (q_1, q_2, \dots, q_n)^T$ [5]. Some of these parameters affect the general position of the object (global rotation and translation), some affect the shape (like a global scaling), and some affect only parts of the object. Since they have a common mathematical formulation as defined in Eqn.(2), we do not need any distinction between these parameters.

Assuming the global deformations \mathbf{T} is differentiable [11], for every point s_i on the surface of the reference model \mathbf{s} , the derivative of s_i with respect to q_i is the Jacobian J_i :

$$J_i = \begin{bmatrix} \vdots & & \vdots \\ \frac{\partial s_i}{\partial q_1} & \dots & \frac{\partial s_i}{\partial q_n} \\ \vdots & & \vdots \end{bmatrix} \quad (3)$$

where each column l of the Jacobian J_i is the gradient of s_i with respect to the parameter q_l . To keep updating the parameters \mathbf{q} to track the global deformations in an expression, we use the following dynamic system updating formula:

$$\dot{\vec{q}} = \vec{f}_g + F_{internal}(\vec{q}) \quad (4)$$

where \vec{f}_g is a n -dimensional displacement called *generalized force* and $F_{internal}(\vec{q})$ is the result of the internal forces of the model (e.g. elasticity). For each frame, the following steps are done iteratively to derive the parameter values at equilibrium:

1. Calculate the 3D external force \vec{f}_i on each point s_i of the deformable model's surface. This force is derived from the 3D displacements between the model point and its closest data point on the face scan.
2. Calculate the generalized force

$$\vec{f}_g = \sum_i \vec{f}_{g_i} = \sum_i J_i^T \vec{f}_i \quad (5)$$

3. Compute the generalized internal force $F_{internal}(\vec{q})$.
4. Calculate the derivative $\dot{\vec{q}}$ as defined in Eqn.(4).
5. Do an Euler integration step:

$$\vec{q} = \vec{q} + \lambda \dot{\vec{q}} \quad (6)$$

where λ is the learning rate.

6. Repeat step 1 to 5, until $\dot{\vec{q}}$ is close to zero.

In our system implementation of the algorithm, we first use an Iterative Closest Point (ICP) method [1] to rigidly align the model and face scans, taking advantage of the dense 3D scan data. Then the face model is divided into several deformable regions whose shape and motion are represented by a few control parameters. Typically, for a smiling expression the face model is divided into 10 small regions with a total of 17 parameters. The changes in these parameters during global tracking are derived from the dynamic system updating scheme described in this section. Because of the small parameter set, the global tracking step is very fast, though it can not capture detailed local deformations.

4. Local deformations

To further recover the local deformations $\mathbf{d}(t)$, as in Eqn. (1), we use a non-rigid 3D shape registration algorithm based on the integration of implicit shape representation and the Free Form Deformations (FFD). In the following description, we refer to the deforming 3D face mesh model as the source surface, and the face range scan as the target surface. We use the 16K fine level mesh model for the local registration.

4.1. The Implicit Shape Representation

Both the 3D face model and the range images, which are surfaces, are implicitly represented in a higher dimensional volumetric space. Given a surface S , the Euclidean distance transform is used to embed this surface as the zero level set of a distance function Φ_S defined in the embedding space Ω :

$$\Phi_S(x, y, z) = \begin{cases} 0 & , (x, y, z) \in S \\ D((x, y, z), S) & , (x, y, z) \in [\Omega - S] \end{cases} \quad (7)$$

where $D((x, y, z), S)$ refers to the min Euclidean distance between the grid location (x, y, z) and the shape S . In shape registration, such a representation facilitates the imposition of constraints on smoothness and coherent correspondence, since one would align the original surfaces as well as their clones that are positioned coherently in the volume plane.

4.2. FFD local registration

To achieve local registration between a source surface S and a target surface D , we aim to recover a deformation field that creates correspondences between the implicit representations Φ_S and Φ_D . We model such a local deformation field $L(\mathbf{x})$, $\mathbf{x} = (x, y, z)$, using a space warping technique, the Cubic B-spline based Free Form Deformations (FFD) [15, 9].

The essence of FFD is to deform an object by manipulating a regular control lattice P overlaid on its volumetric embedding space. When integrated with the implicit shape representation, FFD provides a powerful registration paradigm. We briefly describe the deformations modelled in FFD. Let us consider a lattice of control points,

$$P_{m,n,o} = (P_{m,n,o}^x, P_{m,n,o}^y, P_{m,n,o}^z) \quad (8)$$

where $(m, n, o) \in [1, M] \times [1, N] \times [1, O]$, overlaid to a region $\Gamma = \{\mathbf{x}\} = \{(x, y, z) | 1 \leq x \leq X, 1 \leq y \leq Y, 1 \leq z \leq Z\}$ in the embedding space that encloses the source surface. Suppose the initial configuration of the control lattice P^0 is regular, and the deforming control lattice is $P = P^0 + \delta P$, then in our approach, the local deformation parameters are the incremental FFD parameters, which are the deformations of the control points in all directions:

$$\Theta = \{(\delta P_{m,n,o}^x, \delta P_{m,n,o}^y, \delta P_{m,n,o}^z)\} \quad (9)$$

where $(m, n, o) \in [1, M] \times [1, N] \times [1, O]$. Under these specifications, the deformed position of a point $\mathbf{x} = (x, y, z)$ in the sample domain ¹ given the deformation of

¹We use a narrow band around the zero level set surface as the sample domain to ensure efficiency.

the control lattice from P^0 to P , is defined in terms of a tensor product of Cubic B-splines:

$$\begin{aligned} L(\Theta; \mathbf{x}) &= \mathbf{x} + \delta L(\Theta; \mathbf{x}) \\ &= \sum_{q=0}^3 \sum_{l=0}^3 \sum_{r=0}^3 [B_q(u)B_l(v)B_r(w) \\ &\quad (P_{i+q,j+l,k+r}^0 + \delta P_{i+q,j+l,k+r})] \\ i &= \lfloor \frac{x}{X} \cdot (M-1) \rfloor + 1 \\ j &= \lfloor \frac{y}{Y} \cdot (N-1) \rfloor + 1 \\ k &= \lfloor \frac{z}{Z} \cdot (O-1) \rfloor + 1 \end{aligned} \quad (10)$$

The terms of the deformation component refer to:

- $\delta P_{i+q,j+l,k+r}$, $(q, l, r) \in [0, 3] \times [0, 3] \times [0, 3]$ are the deformations of point \mathbf{x} 's 64 adjacent control points.
- $B_q(u)$ is the q^{th} , $B_l(v)$ is the l^{th} and $B_r(w)$ is the r^{th} basis function of a Cubic B-spline.
- $\delta L(\mathbf{x}) = \sum_{q=0}^3 \sum_{l=0}^3 \sum_{r=0}^3 B_q(u)B_l(v)B_r(w) \delta P_{i+q,j+l,k+r}$ is the incremental deformation for point \mathbf{x} .

Having defined the form of the local deformation field $L(\mathbf{x})$ with respect to the FFD parameters $\Theta = \delta P$, local registration is now equivalent to finding the control lattice deformation δP such that the deformed source surface coincides with the target surface. The Sum-of-Squared-Differences (SSD) criterion is used as a data-driven term to recover the parameters:

$$E_{data}(\Theta) = \iiint_{\Omega} (\Phi_D(\mathbf{x}) - \Phi_S(L(\Theta; \mathbf{x})))^2 d\mathbf{x} \quad (11)$$

In order to further preserve the regularity of the recovered registration flow, one can consider a smoothness term on the local deformation field δL . We consider a computationally efficient smoothness term:

$$E_{smooth}(\Theta) = \iiint_{\Omega} \left(\left\| \frac{\partial}{\partial x} \delta L(\Theta; \mathbf{x}) \right\|^2 + \left\| \frac{\partial}{\partial y} \delta L(\Theta; \mathbf{x}) \right\|^2 + \left\| \frac{\partial}{\partial z} \delta L(\Theta; \mathbf{x}) \right\|^2 \right) d\mathbf{x} \quad (12)$$

An additional implicit smoothness constraint is also imposed by the B-Spline FFD, which guarantees C^1 continuity at control points and C^2 continuity everywhere else.

We can also enhance the accuracy of the tracking system by imposing correspondence constraints on certain feature points such as tip of the nose and corners of the eyes. Assuming we have n_c features, and for each of them, there is a pair of corresponding points, \mathbf{x}_{si} on the source surface S and \mathbf{x}_{di} on the target surface D , where $i = 1, \dots, n_c$. Then the feature correspondence constraints can be expressed as

$$E_{feature}(\Theta) = \sum_i (L(\Theta; \mathbf{x}_{si}) - \mathbf{x}_{di})^2; i \in [1, n_c] \quad (13)$$

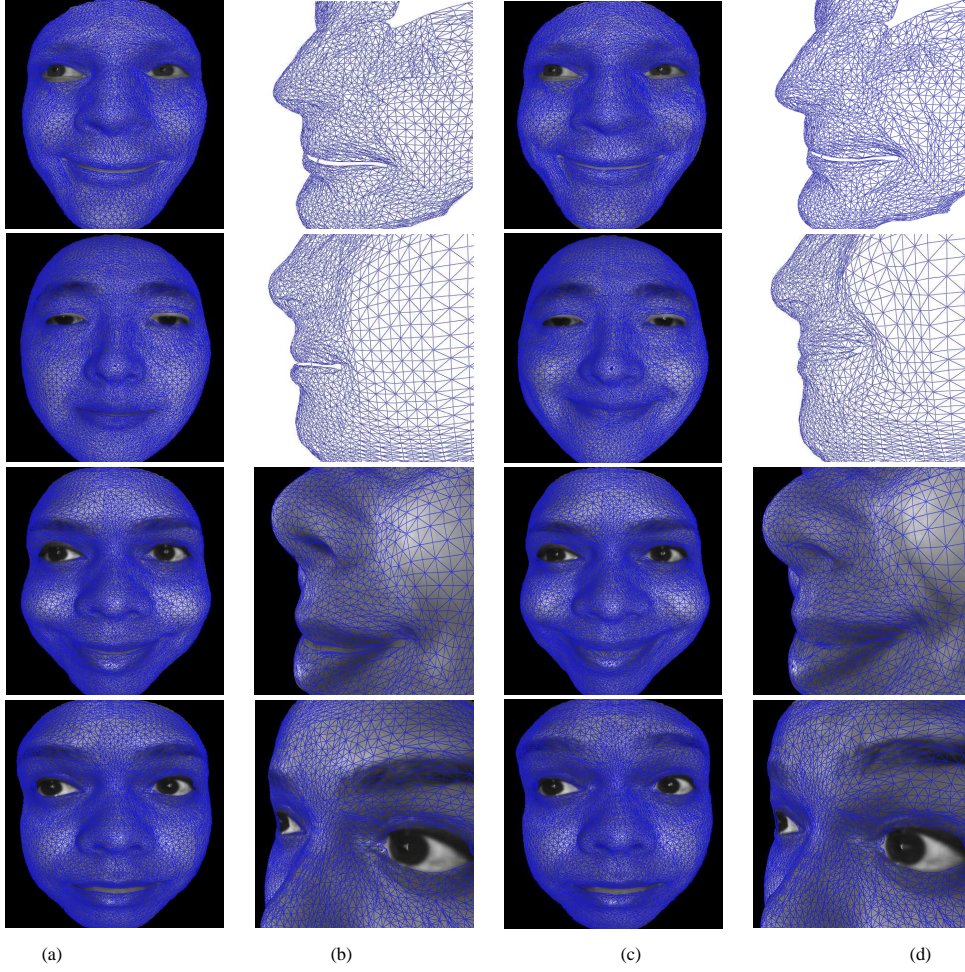


Figure 2. [Top Row]: Snapshots of the *smile* expression of subject 1. [Second Row]: The *smile* expression of subject 2. [Third Row]: The *smile* expression of subject 3. [Bottom Row]: The *Raising eyebrow* expression of subject 3. [Column a]: Front view of frame 1. [Column b]: Close-up view of Column a (without range scan - for showing details; with range scan - for showing correspondences). [Column c]: Front view of frame 2. [Column d]: Close-up view of Column c.

The data-driven term, the smoothness constraint term, and the feature correspondence constraint term can be integrated in a single objective function,

$$\begin{aligned}
E(\Theta) &= E_{data}(\Theta) + \alpha E_{smooth}(\Theta) + \beta E_{feature}(\Theta) \\
&= \iint \iint_{\Omega} (\Phi_D(\mathbf{x}) - \Phi_S(L(\Theta; \mathbf{x})))^2 d\mathbf{x} \\
&+ \alpha \iint \iint_{\Omega} \left(\left\| \frac{\partial}{\partial x} \delta L(\Theta; \mathbf{x}) \right\|^2 + \left\| \frac{\partial}{\partial y} \delta L(\Theta; \mathbf{x}) \right\|^2 \right. \\
&\quad \left. + \left\| \frac{\partial}{\partial z} \delta L(\Theta; \mathbf{x}) \right\|^2 \right) d\mathbf{x} \\
&+ \beta \sum_i (L(\Theta; \mathbf{x}_{si}) - \mathbf{x}_{di})^2
\end{aligned} \tag{14}$$

where $i \in [1, n_c]$, n_c is the number of feature points, and α and β are the constants balancing the contributions from different terms. Using the calculus of variations and a gra-

dient descent method, such an objective function can be optimized to recover the deformation parameters Θ ,

$$\begin{aligned}
\frac{\partial E(\Theta)}{\partial \theta_i} &= -2 \iint \iint_{\Omega} [(\Phi_D(\mathbf{x}) - \Phi_S(L(\Theta; \mathbf{x}))) \\
&\quad \cdot \nabla \Phi_S(L(\Theta; \mathbf{x})) \cdot \frac{\partial \delta L(\Theta; \mathbf{x})}{\partial \theta_i}] d\mathbf{x} \\
&+ 2\alpha \iint \iint_{\Omega} \left[\frac{\partial}{\partial x} \delta L(\Theta; \mathbf{x}) \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial x} \delta L(\Theta; \mathbf{x}) \right) \right. \\
&\quad + \frac{\partial}{\partial y} \delta L(\Theta; \mathbf{x}) \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial y} \delta L(\Theta; \mathbf{x}) \right) \\
&\quad \left. + \frac{\partial}{\partial z} \delta L(\Theta; \mathbf{x}) \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial z} \delta L(\Theta; \mathbf{x}) \right) \right] d\mathbf{x} \\
&+ 2\beta \sum_i [(L(\Theta; \mathbf{x}_{si}) - \mathbf{x}_{di}) \frac{\partial}{\partial \theta_i} (L(\Theta; \mathbf{x}_{si}))]
\end{aligned} \tag{15}$$

and consequently the local registration field $L(\Theta; \mathbf{x})$. Correspondences thus can be established between points $\mathbf{x} = (x, y, z)$ on the source surface and the points $L(\mathbf{x})$ on the target surface. And the displacements between the corre-

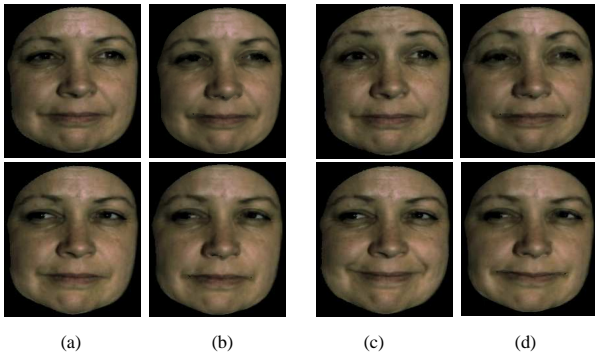


Figure 3. [Top Row]: Comparison between original texture of a subject’s colored range scans and synthesized texture of the tracking face control mesh, for the *raising eyebrow* expression. [Second row]: Comparison for the *smile* expression. (a) Snapshot 1 from the original scan data. (b) Snapshot 1 from the synthesized rendering of the tracking result. (c) Snapshot 2 from the original scan data. (d) Snapshot 2 from the synthesized rendering of the tracking result.

sponding points consist of the local deformations $\mathbf{d}(t)$ in Eqn. (1).

5. Experimental Results

We conducted tracking experiments using the dynamic facial scans of five different subjects. These data are acquired using the 3D high resolution shape acquisition system described by [19]. For each subject, data for two different expressions are collected: the *smile* expression and the *raising eyebrow* expression. On each data sequence, we first register its first frame with the face model (at fine level 16K nodes), then we keep tracking the intra-frame deformations using the tightly coupled global and local tracking algorithm. This hierarchical tracking protocol results in efficient and very detailed fitting to the 3D face scan data. Example tracking results are shown in Figure 2. The fine details in an expression captured using our method is demonstrated in Figure 2(a-b), and the high accuracy of the intra-frame correspondences established during tracking is demonstrated in Figure 2(c-d).

Our system is implemented using C++ under the Linux environment. All our experiments run at interactive rate on a Pentium Xeon 3GHz dual processor platform.

For qualitative evaluation of the model point tracking results, we compare the texture of the original scan data with the synthesized texture based on the tracking results. We generate synthetic texture for the 16K-node face control mesh that tracks the range scans, by applying the texture retrieved from the registered first frame to the remaining

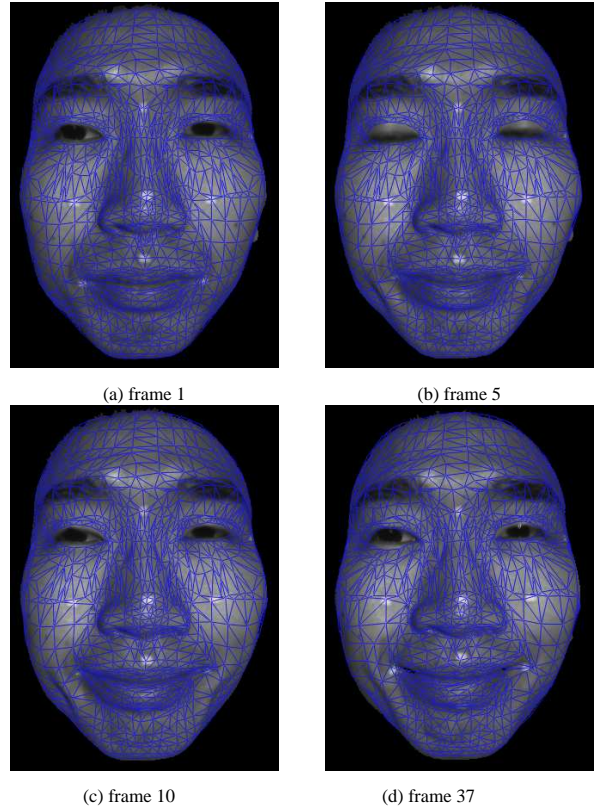


Figure 4. Selected tracking results of a ‘smile’ sequence, with 50 frames in total. The resulting meshes are illustrated in blue color and white dots are attached markers for verification purposes only. (a) frame 1, (b) frame 5, (c) frame 10, and (d) frame 37.

frames of each sequence, adding in shading and shadow effects considering the change in facial geometry during the expression. The comparison results on frames from example tracking sequences are shown in Figure 3. Ideally, we should have compared the reflectance maps of the original scans with the synthesized reflectance maps. But our comparison of direct luminance-based textures provides sufficient evidence on the validity of the tracking result.

To further conduct quantitative validation on the accuracy of the tracking algorithm, we perform a number of experiments on 3D facial expression sequences with attached markers. The markers are for validation purpose only and are not used for the tracking. In order to be detected successfully, the size of markers is around 4mm by 4mm. An example tracking result is demonstrated in Figure 4, where the blue meshes are resulting coarse level control mesh, and the white dots are the attached markers for verification purposes only. Figures 5-8 show the algorithm’s tracking error estimations on the mouth corner, upper mouth, cheek, and nose tip respectively. As we can see, in most cases the tracking error in 3D is around 1mm. This error is very low given

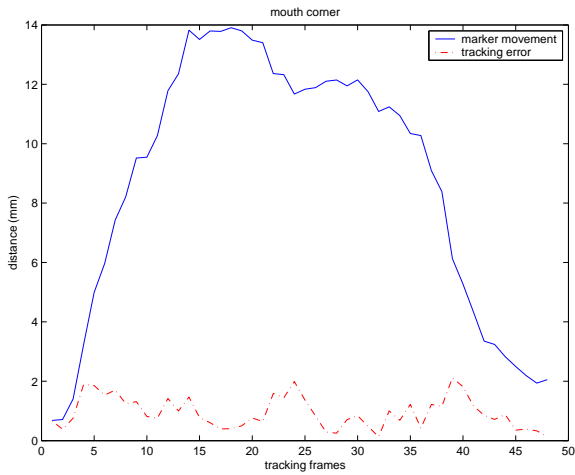


Figure 5. Tracking error of the marker on a mouth corner.

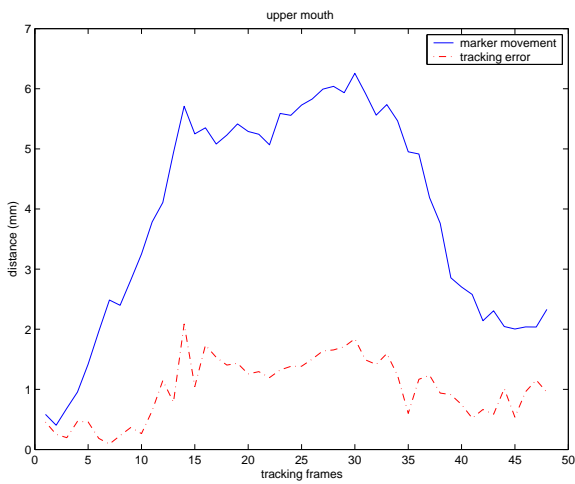


Figure 6. Tracking error of the marker on the upper mouth.

that the resolution of the 3D range scan data is $0.5mm$ in X and Y directions and $0.2mm$ in Z direction.

6 Conclusions

In this paper, we have developed a hierarchical facial expression tracking framework for establishing intra-frame correspondences among high quality 3D dynamic facial range scan data. The 3D face range scan data are acquired at speeds that exceed regular video frame rate. Based on our proposed hierarchical tracking framework, we are able to efficiently parameterize such a large amount of data by dealing with both large-scale deformations and free-form style fine-details existing in the facial expressions. The accuracy and resolution of our method allow us to capture and track subtle expression details and hence to use the tracking

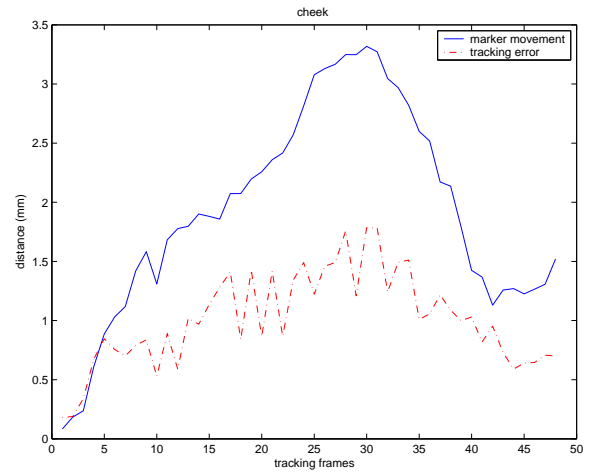


Figure 7. Tracking error of the marker on a cheek.

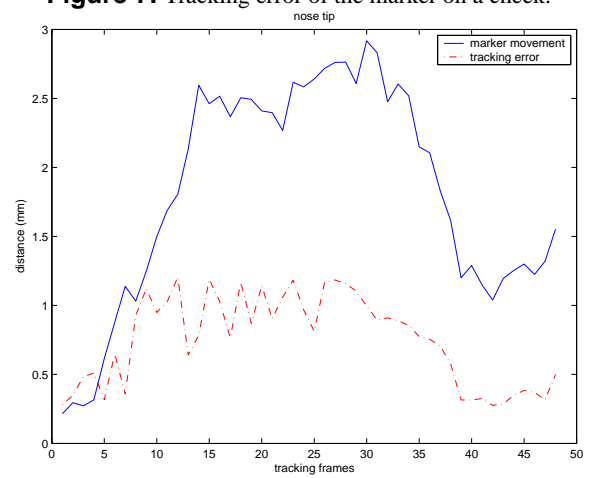


Figure 8. Tracking error of the marker on the nose tip.

parameters for motion analysis and expression recognition.

In our future work, we will address the problem of automated robust detection of facial features and feature correspondences to constrain the global and local deformations of the face model, thus to further decrease the tracking error. We will also look into the incorporation of skin reflectance modelling for robust tracking under different illumination conditions, considering both geometry and texture information.

Acknowledgements

The authors are grateful to Prof. Peisen Huang in the Mechanical Engineering department at the State University of New York, Stony Brook, for the help on 3D dynamic facial expression data acquisition. The authors also want to thank S. Goldenstein and C. Vogler for their help with the tracking system[5] and S. Venkataraman, Z. Li for initial face mesh work. Support for this research was provided by

NSF grants ACI-0313184 and ITR-0205671.

References

- [1] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 1992.
- [2] J. Davis, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. In *CVPR'03*, pages 359–366, June 2003.
- [3] D. DeCarlo and D. Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.
- [4] L. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *CVPR'94*, pages 76–83, 1994.
- [5] S. K. Goldenstein, C. Vogler, and D. Metaxas. Statistical cue integration in dag deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):801–813, 2003.
- [6] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making faces. In *SIGGRAPH'98*, pages 55–66, 1998.
- [7] P. S. Huang, C. Zhang, and F. P. Chiang. High-speed 3-d shape measurement based on digital fringe projection. *Opt. Eng.*, 42(1):163–168, 2003.
- [8] X. Huang, N. Paragios, and D. Metaxas. Establishing local correspondences towards compact representations of anatomical structures. In *MICCAI'03*, pages 926–934, 2003.
- [9] S. Lee, K.-Y. Chwa, J. Hahn, and S. Y. Shin. Image morphing using deformation techniques. *Journal of Visualization and Computer Animation*, 7:3–23, 1996.
- [10] J. Lien, T. Kanade, A. Zlochow, J. Cohn, and C. Li. Subtly different facial expression recognition and expression intensity estimation. In *CVPR'98*, pages 853–859, 1998.
- [11] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.
- [12] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.
- [13] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 8:712–721, 1999.
- [14] S. Rusinkiewicz, O. Hall-Holt, and L. Marc. Real-time 3d model acquisition. In *SIGGRAPH'02*, volume 1281 of *3D acquisition and image based rendering*, pages 438 – 446, 2002.
- [15] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *SIGGRAPH'86*, pages 151–160, Aug. 1986.
- [16] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. In *CVPR'94*, pages 70–75, 1994.
- [17] J. yong Noh and U. Neumann. Expression cloning. In *SIGGRAPH'01*, pages 277–288, 2001.
- [18] L. Zhang, B. Curless, and S. M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *CVPR'03*, pages 367–374, June 2003.
- [19] S. Zhang and P. S. Huang. High-resolution, real-time 3-d shape acquisition. In *IEEE Workshop on Real-time 3D Sensors and Their Use (joint with CVPR'04)*, 2004.
- [20] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.