Dense 3D Motion Capture for Human Faces

Yasutaka Furukawa
University of Washington, Seattle, USA
furukawa@cs.washington.edu

Jean Ponce *
Ecole Normale Supérieure, Paris, France
Jean.Ponce@ens.fr

Abstract

This paper proposes a novel approach to motion capture from multiple, synchronized video streams, specifically aimed at recording dense and accurate models of the structure and motion of highly deformable surfaces such as skin, that stretches, shrinks, and shears in the midst of normal facial expressions. Solving this problem is a key step toward effective performance capture for the entertainment industry, but progress so far has been hampered by the lack of appropriate local motion and smoothness models. The main technical contribution of this paper is a novel approach to regularization adapted to nonrigid tangential deformations. Concretely, we estimate the nonrigid deformation parameters at each vertex of a surface mesh, smooth them over a local neighborhood for robustness, and use them to regularize the tangential motion estimation. To demonstrate the power of the proposed approach, we have integrated it into our previous work for markerless motion capture [9], and compared the performances of the original and new algorithms on three extremely challenging face datasets that include highly nonrigid skin deformations, wrinkles, and quickly changing expressions. Additional experiments with a dataset featuring fast-moving cloth with complex and evolving fold structures demonstrate that the adaptability of the proposed regularization scheme to nonrigid tangential motion does not hamper its robustness, since it successfully recovers the shape and motion of the cloth without overfitting it despite the absence of stretch or shear in this case.

1. Introduction

The most popular approach to motion capture today is to attach reflective markers to the body and/or face of an actor, and track these markers in images acquired by multiple calibrated video cameras [3]. The marker tracks are then matched, and triangulation is used to reconstruct the corresponding position and velocity information. The accuracy of any motion capture system is limited by the temporal and spatial resolution of the cameras, and the number of reflective markers to be tracked, since matching becomes difficult with too many markers that all look alike. On the other hand, although relatively few (say, 50) markers may be sufficient to recover skeletal body configurations, thousands (or even more) may be needed to accurately recover the complex changes in the fold structure of cloth during body motions [23], or model subtle facial motions and skin deformations [4, 9, 16, 17]. Computer vision methods for markerless motion capture (possibly assisted by special make-up or random texture patterns painted on a subject) offer an attractive alternative, since they can (in principle) exploit the dynamic texture of the observed surfaces themselves to provide reconstructions with fine surface details and dense estimates of nonrigid motion. Such a technology is indeed emerging in the entertainment and medical industries [1, 2]. Several approaches to local *scene flow* estimation have also been proposed in the computer vision literature to handle less constrained settings [5, 13, 15, 18, 20, 21], and recent research has demonstrated the recovery of dense human body motion using shape priors or pre-acquired laser-scanned models [6, 22]. Despite this progress, a major impediment to the deployment of facial motion capture technology in the entertainment industry is its inability (so far) to capture fine expression detail in certain crucial areas such as the mouth, which is exacerbated by the fact that people are very good at picking unnatural motions and "wooden" expressions in animated characters. Therefore, complex facial expressions remain a challenge for existing approaches to motion capture, because skin stretches, shrinks, and shears much more than other materials such as cloth or paper, and the local motion models typically used in motion capture are not adapted to such deformations. The main technical contribution of this paper is a novel approach to regularization specifically designed for nonrigid tangential deformations via a local linear model. It is simple but, as shown by our experiments, very effective in capturing extremely complicated facial expressions.

*Willow Project-Team, Laboratoire d'Informatique de l'Ecole Normale Supérieure, ENS/INRIA/CNRS UMR 8548

1.1. Related Work

Three-dimensional *active appearance models* (AAMs) are often used for facial motion capture [12, 14]. In this approach, parametric models encoding both facial shape and appearance are fitted to one or several image sequences. AAMs require an a priori parametric face model and are, by design, aimed at tracking relatively coarse facial motions rather than recovering fine surface detail and subtle expressions. *Active sensing* approaches to motion capture use a projected pattern to independently estimate the scene structure in each frame, then use optical flow and/or surface matches between adjacent frames to recover the three-dimensional motion field, or *scene flow* [10, 24]. Although qualitative results are impressive, these methods typically do not exploit the redundancy of the spatio-temporal information, and may be susceptible to error accumulation over time due to the concatenation of local motion fields [19]. In addition the estimated motion may be erroneous because the projected patterns typically make accurate tangential tracking difficult. Several *passive* approaches to scene flow computation have also been proposed [5, 13, 15, 18, 21]. However, these approaches suffer from two limitations: First, they have so far mostly been restricted to simple motions with little occlusion. The second limitation is again accumulating drift. We have recently proposed a mesh-based motion capture algorithm [9] that does not suffer from accumulation errors, and handles complicated surface deformation. However, it assumes *locally rigid* motion and is not designed for nonrigid deformations with much stretching, shrinking or shearing, such as those common in facial expressions. In general, accurate facial motion capture remains an unsolved challenge for existing approaches to motion capture. First, many algorithms focus more on good visualization than accurate motion recovery. This makes sense in cases such as full-body motion capture, where clothes may not have enough texture to yield high-resolution motion and, on the other hand, cloth animation is often visually plausible even when the motion is not physically accurate. The situation is very different in facial motion capture, since people are, as noted earlier, very good at picking unnatural expressions. Second, motion-capture algorithms are often simply not designed for handling nonrigid tangential motions. For example, a locally rigid motion model, although perfectly acceptable for capturing the motion of paper and cloth, may smooth out all the details of a facial expression. The algorithm proposed in [4] captures fine-scale facial geometry and motion, but it focuses mostly on the plausible synthesis of expression wrinkles. It also requires a user to apply paint on a face at expected wrinkle locations before-hand, which is time consuming and may not work for unexpected facial expressions (see Fig. 3 for example, with wrinkles on a person's neck).

The challenge in our work is the development of a smart regularization term that allows severe nonrigid deformation but is also robust especially where texture information becomes unreliable due to fast motion, self-occlusions, poor image texture, etc. The Laplacian operator used for regularization by several current algorithms [6, 9, 15, 18] is too weak to handle complicated surface deformations in challenging sequences such as those shown in Fig. 5. A tangential rigidity constraint has been shown to be very effective in such cases [9], but it does not work well with intricate facial expressions whose deformation contains a lot of stretch, shrink and shear. Our solution to this problem is to model and estimate in a stable fashion the tangential nonrigid deformation. More concretely, given a mesh model in a certain frame, we first estimate the tangential nonrigid deformation at each vertex by projecting its neighboring vertices onto the tangent plane and computing a 2D linear transformation that maps the projected vertices from the reference frame to the current one. Second, we smooth these deformation parameters over a local neighborhood for robustness, which is especially important in surface areas with unreliable image information (see Fig. 6 for the effects of smoothing). The estimated nonrigid deformation is then used to define a novel adaptive tangential rigidity term. Our method is very simple yet works well in various challenging cases. In reality, of course, the skin has a complicated layered structure, and its physical behaviour results from the interaction between those layers, but a simple per-vertex linear deformation model has been proven effective in our experiments.

To demonstrate the power of the proposed approach, we have integrated it into our previous work for markerless motion capture [9], dubbed *FP08* in the rest of this presentation. We have tested our implementation on three real face datasets with complicated, fast-changing expressions, and show in Section 4 that it successfully and accurately captures intricate facial details in each case. Additional experiments with a dataset featuring fast-moving cloth with complex and evolving fold structures demonstrate that the adaptability of the proposed regularization scheme to nonrigid tangential motion does not hamper its generality or robustness, since it successfully recovers the shape and motion of the cloth without overfitting it despite the absence of stretch or shear in this case. We compare in Section 4 our results with those obtained by the original FP08 algorithm, and also perform some qualitative evaluations to show the effects of the key components in our algorithm. The rest of the article is organized as follows. Section 2 briefly reviews the FP08 algorithm proposed in [9] for completeness. Section 3 explains how to model and estimate tangential nonrigidity, then use it in the motion capture algorithm, which is the main contribution of the paper. We present our experimental results in Sect. 4, then conclude the paper with a discussion of future work in Sect. 5.

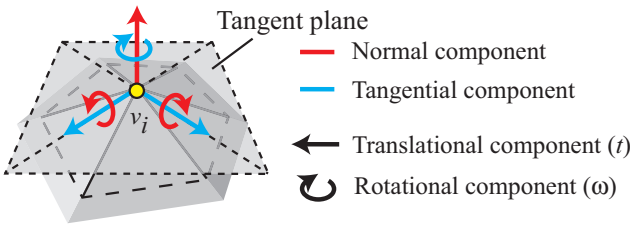

Figure 1. The local rigid motion can be decomposed into the *tangential* and *normal* components (reproduced with permission from [9]). In this paper, we also model nonrigid surface deformation in the tangent plane from the reference frame to control tangential rigidity of a surface such as stretch, shrink, and shear.

2. The FP08 Algorithm

We briefly review the algorithm proposed in [9] in this section. The instantaneous geometry of the observed scene is represented by a polyhedral mesh with fixed topology. An initial mesh is constructed in the first frame by using the publicly available PMVS software for multi-view stereo (MVS) [8] and Poisson surface reconstruction software [11] for meshing, then its deformation is captured by tracking its vertices $\{v_1, \ldots, v_n\}$ over time. The goal of the algorithm is to estimate in each frame f the position v_i^f of each vertex v_i (from now on, v_i^f will be used to denote both the vertex and its position). Note that each vertex may or may not be tracked at a given frame, including the first one, allowing the system to handle occlusion, fast motion, and parts of the surface that are not visible initially. The three steps of the tracking algorithm –local motion estimation, global surface deformation, and filtering– are detailed in the following sections.

2.1. Local Rigid Motion Estimation

At each frame, the FP08 algorithm approximates a local surface region around each vertex by its tangent plane, and estimates the corresponding local 3D rigid motion with six degrees of freedom. The algorithm uses two techniques to improve robustness and accuracy. The first one is motion decomposition: As illustrated by Fig. 1, among six degrees of freedom, three parameters encode *structure* or *normal* information (depth and surface normal), while the remaining three contain *tangential* motion information (translation in the tangent plane and rotation about the surface normal). Instead of directly estimating all six parameters from the beginning, which is susceptible to local minima, the normal parameters are first found by optimizing a *structure photometric consistency* function, then all the six parameters are refined by optimizing a *motion photometric consistency* function. The second key to robustness is an expansion strategy that makes use of the spatial consistency of local motion information.

2.2. Global Surface Deformation

Based on the estimated local motion parameters, the whole mesh is then deformed by minimizing the sum of three energy terms:

$$\sum_i |v_i^f - \hat{v}_i^f|^2 + \eta_1 |[\zeta_2 \Delta^2 - \zeta_1 \Delta] v_i^f|^2 + \eta_2 E_r(v_i^f). \quad (1)$$

The first *data* term simply measures the squared distance between the vertex position v_i^f and the position $\hat{v}_i^f$ estimated by the local estimation process. The second term uses the (discrete) Laplacian operator Δ of a local parameterization of the surface in v_i to enforce smoothness [7] (the values $\zeta_1 = 0.6$ and $\zeta_2 = 0.4$ are used in all the experiments of [9] and in the present paper as well). This term is very similar to the Laplacian regularizer used in many other algorithms [6, 15, 18]. The third term is also for regularization, and it enforces (local) tangential rigidity with no stretch, shrink or shear. The total energy is minimized with respect to the 3D positions of all the vertices by a conjugate gradient method.

2.3. Filtering Out Erroneous Local Motion

After surface deformation, the residuals of the data and tangential rigidity terms are used to filter out erroneous motion estimates. Concretely, these values are first smoothed, and a (smoothed) local motion estimate is deemed an outlier if at least one of the two residuals exceeds a given threshold. The three steps are iterated a couple of times to complete tracking in each frame, the local motion estimation step only being applied to vertices whose parameters have not already been estimated or filtered out. Please see [9] for more details of the algorithm.

2.4. Adapting FP08

In addition to the new tangential rigidity term explained in the next section, we have made two (minor) modifications to the local rigid motion estimation step (Sect. 2.1) mainly to improve the visual quality of reconstructed meshes. First, we have observed that the surface obtained after motion optimization is often noisier than the one obtained from structure optimization. This is probably because the shading and shadows of an object might change from frame to frame, making some of the texture information unreliable in the motion estimation step where different frames must be compared. Therefore, we perform the structure optimization once again after the motion optimization to refine the structure parameters while fixing the remaining motion parameters (see [18] for a similar procedure). The second modification is the removal of an error term in the local structure and motion optimization, which penalizes the deviation of the parameters from their initial guesses. We have observed that the proposed system is stable without such a

term that may simply add bias to the data information. Although differences resulting from these two modifications are small, their effects on noise reduction is noticeable in certain places. [1]

3. A New Regularization Scheme

As mentioned before and shown in our experiments later, the tangential rigidity constraint in Eq. (1) is too strict for facial motion capture since it does not allow skin deformations including stretch, shrink and shear. Regularizing the tangential motion is, on the other hand, a key factor in handling complicated surface deformations (see Fig. 5 for examples). Thus, instead of assuming static edge lengths as in [9], we propose in this paper to estimate the nonrigid tangential deformation from the reference frame to the current one at each vertex, and use that information to compute target edge lengths. The estimation of the tangential deformation is performed at each frame before starting the motion estimation, and the parameters are fixed within a frame. The actual estimation consists of two steps –independent estimation at each vertex, and smoothing over local surface neighborhood– that are detailed in the next sections.

3.1. Estimating Nonrigid Surface Deformation

We approximate the nonrigid tangential surface deformation from the reference frame to the current one by a 2D linear transformation in the tangent plane of each vertex (the origins of the corresponding coordinate frames are aligned, avoiding the need for a translation term). Concretely, given a vertex v_i^f at frame f, the adjacent vertices are first projected onto the tangent plane at v_i^f (Fig. 2, left). We attach an arbitrary 2D coordinate frame to the tangent plane by aligning its origin with v_i^f, and use $x_i^f(j)$ to denote the position of the projection of each neighbor v_j^f in this coordinate frame. After performing the same projection procedure at the reference frame f_0, we solve for a linear deformation A_i^f that maps $x_i^{f_0}(j)$ onto $x_i^f(j)$ for every adjacent vertex v_j in $\mathbf{N}(v_i)$:

$$x_i^f(j) = A_i^f x_i^{f_0}(j).$$

Here, A_i^f is a 2×2 matrix, $x_i^f(j)$ is a vector in $\mathbb{R}^2$, and the above equation adds two constraints for each neighbor. Since each vertex has at least two (and typically more) neighbors, we compute A_i^f by solving a linear least squares problem.

3.2. Smoothing Nonrigid Deformation Parameters

The second step is to smooth the nonrigid deformation parameters over the surface for robustness, based on the assumption that the nonrigid surface deformation is spatially smooth, and nearby vertices follow similar deformations. [2] More concretely, we smooth nonrigid deformation parameters A_i^f over the surface instead of allowing each vertex to have independent values. However, the deformation parameters for adjacent vertices are expressed in different coordinate frames attached to different tangent planes, and we thus need to align these coordinate frames. Given a pair of adjacent vertices v_i^f and v_j^f in frame f, we simply assume that their tangent planes are identical, and first estimate the 2D rotation matrix R_{ij}^f that aligns the vectors $x_i^f(j) - x_i^f(i)$ with $x_j^f(j) - x_j^f(i)$, then the translation vector t_{ij}^f that maps $x_i^f(i)$ onto $x_j^f(i)$ (Fig. 2, center). Note that we are not estimating a deformation but simply aligning coordinate frames, and just need a 2D rigid transformation (rotation and translation). Of course, the registration is not perfect but, again, this is not a critical issue. Assuming that nonrigid tangential deformation is consistent between adjacent vertices, we expect the following equations to hold for any 2D point x:

$$R_{ij}^f(A_i^f x) + t_{ij}^f = A_j^f(R_{ij}^{f_0} x + t_{ij}^{f_0}).$$

The left side of this equation characterizes the position x of a point that first follows the deformation around vertex v_i at the reference frame f_0, and is then mapped onto the other coordinate frame at frame f. Its right side characterizes the position x of a point that is first mapped onto the second coordinate frame at the reference frame f_0, then follows the deformation about vertex v_j (Fig. 2, right). This equation can be rewritten as

$$(R_{ij}^f A_i^f - A_j^f R_{ij}^{f_0})x = A_j^f t_{ij}^{f_0} - t_{ij}^f,$$

and since it should hold for all x, and $A_j^f t_{ij}^{f_0} - t_{ij}^f$ should be very close to 0 by construction, we obtain the (approximate) constraint

$$A_i^f = R_{ij}^{fT} A_j^f R_{ij}^{f_0}.$$

This relation is finally used to smooth each vertex by repeating 8 times the following local averaging operation:

$$A_i^f \leftarrow \frac{1}{1+|\mathbf{N}(v_i)|}[A_i^f + \sum_{v_j \in \mathbf{N}(v_i)} R^f T_{ij} A_j^f R_{ij}^{f_0}].$$

3.3. Adaptive Tangential Rigidity Term

Given a vertex v_i^f and its nonrigid deformation parameters A_i^f at frame f, the (3D) length e_{ij}^f of an edge between v_i^f and its neighbor $v_j^f (v_j \in \mathbf{N}(v_i))$ should be

$$\hat{e}_{ij}^f = e_{ij}^{f_0} \frac{|A_i^f x_i^{f_0}(j)|}{|x_i^{f_0}(j)|}, \quad (2)$$

[1] See videos on our project website `http://www.cs.washington.edu/homes/furukawa`.

[2] The assumption is reasonable in many cases where external forces to the surface stem from a few locations, yielding locally consistent nonrigid deformations, e.g., facial expressions governed by a few active muscles.

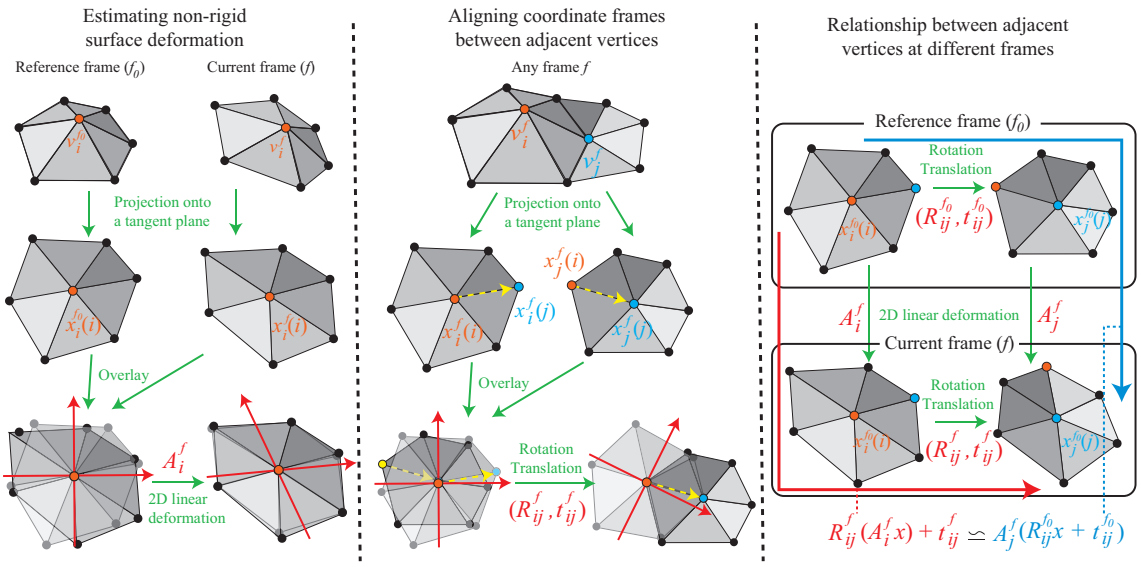

Figure 2. We approximate the nonrigid deformation around each vertex by a 2D linear transformation in its tangent plane. Left: estimation of the deformation parameters from the reference frame f_0 to the current one f. Center: alignment of different coordinate frames between neighboring vertices. Right: the relationship between adjacent vertices in two different frames, which is used to smooth deformation parameters.

where $e_{ij}^{f_0}$ is the original (3D) edge length in the reference frame f_0, and the rest of the term measures the amount of stretch and shrink from frame f_0 to f. (Here, as usual, we have assumed that local coordinate system was centered in v_i^f). Thus, our tangential rigidity term $E_r(v_i^f)$ for a vertex v_i^f in the global mesh deformation step (1) is given by

$$\sum_{v_j \in \mathbf{N}(v_i)} \max[0, (e_{ij}^f - \hat{e}_{ij}^f)^2 - \tau^2], \tag{3}$$

which is the sum of squared differences between the actual edge lengths and those predicted by Eq. (2). The term τ is used to make the penalty zero when the deviation is small so that this regularization term is enforced only when the data term is unreliable and the error is large. In all our experiments, τ is set to be 0.2 times the average edge length of the mesh at the first frame.

4. Experimental Results

We have implemented the proposed method and tested it using three real face sequences (*face1*, *face2* and *face3*) kindly provided by Image Movers Digital and one cloth sequence (*pants*), kindly provided by R. White, K. Crane and D.A. Forsyth [23]. In each case, the data consists of image streams from multiple synchronized and calibrated cameras. Sample input images are shown in Fig. 3, and Table 1 provides some characteristics and choices of parameters for each dataset. Note that all the other parameters are fixed and the same for all the datasets. The *pants* and *face1* videos contain fast and complex motions but without much stretch nor shrink, and the *face2* and *face3* sequences contain complicated facial expressions with highly nonrigid deformations, where an accurate estimation of tangential deformations is necessary for successful motion capture.

Table 1. Characteristics of the datasets. N_v, N_c, N_f, N_p, T, η_1 and η_2 respectively denote the number of vertices in a mesh, the number of cameras, the number of frames, the number of effective pixels (an object appears small in some datasets), an average running time of the algorithm per frame in minutes, and weights associated with two regularization terms in (1).

	N_v	N_c	N_f	N_p	T	η_1	η_2
pants	8652	8	173	0.2M	0.42	10	10
face1	39612	10	325	0.3M	1.6	5	10
face2	75603	10	400	0.3M	2.2	5	10
face3	75603	10	430	0.3M	2.1	5	10

As stated in our previous paper [9], which is the basis of our implementation, the publicly available PMVS software [8] and a meshing software [11] are used to initialize a mesh model in the first frame. For the three face datasets, we have manually added a hole at the mouth to the meshe, since its topology is fixed in FP08. All the algorithms are implemented in C++ and a dual quad-core 2.66GHz linux machine has been used for the experiments.

Figure 3 shows, for each dataset, a sample input image, a reconstructed mesh model, the estimated motion, and a texture-mapped model for two frames with interesting structure and/or motion. [3] The motion information at

[3] See our project website for videos `http://www.cs.washington.edu/homes/furukawa`.

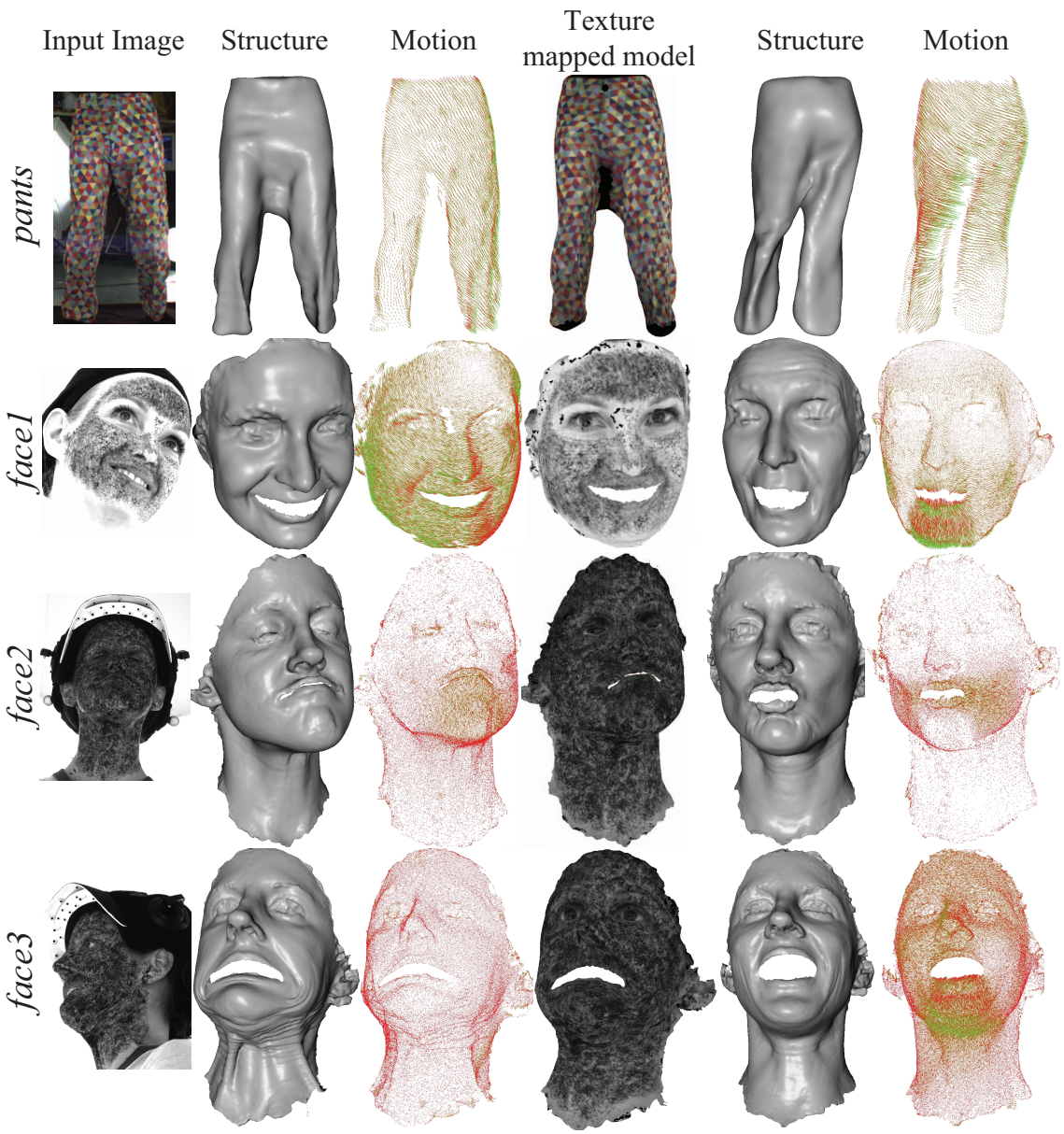

Figure 3. From left to right, a sample input image, reconstructed mesh model, estimated motion, and a texture mapped model for one frame with interesting structure/motion for each dataset. The right two columns show the results in another interesting frame. See text for details.

each vertex is illustrated by a colored line segment that connects its 3D locations from the previous frame (red) to the current (green). Textures are mapped onto the mesh by averaging the back-projected textures from every visible image in every tracked frame as in [9]. This is an effective method for qualitative assessment, since the texture will only appear sharp when the estimated structure and motion information are accurate throughout the sequence. As shown by the figure, our algorithm successfully recovers various facial structure and deformation including highly nonrigid skin deformation with complicated wrinkles at the neck, cheeks, and lips. The computed model textures also appear sharp excluding exceptional places such as eyes for face datasets and the inner thigh region for the *pants* dataset, where tracking is very difficult. The *pants* videos form an interesting dataset for our algorithm in two respects: First, since the cloth does not stretch nor shrink much, tangential deformations needs not be considered, and one may fear that our approach will overfit the deformations and create unnecessary wrinkles. A shown by Figure 3, this is not the case, and our algorithm successfully captures accurate surface deformation, demonstrating the robustness of the system. Second, due to occlusions between inner thighs, the initial mesh model is not accurate there, causing tracking problems for

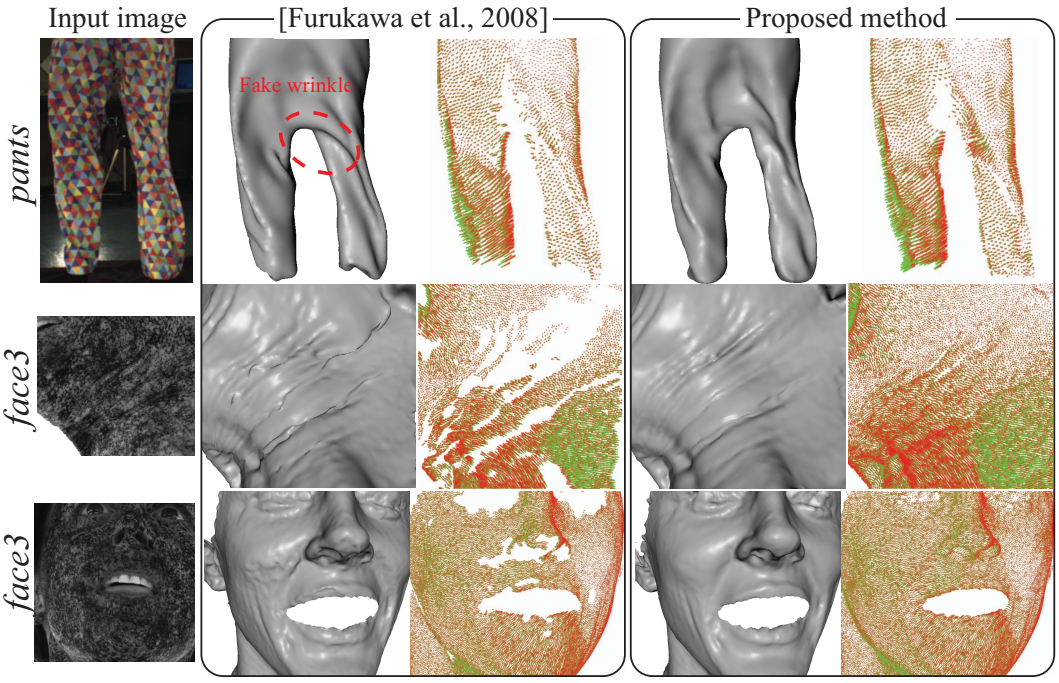

Figure 4. Comparison of the proposed algorithm with FP08 [9]. The proposed algorithm can handle highly nonrigid surface deformations as well as surface regions with inaccurate mesh initialization. See texts for details.

FP08 [9] and yielding fake wrinkles due to the strong rigidity constraint, whereas the use of our adaptive tangential rigidity term avoids such artifacts (top of Fig. 4). Figure 4 shows qualitative comparisons between the proposed algorithm and FP08, illustrating (as expected, since it is designed for surfaces that bend but don't stretch or shear) that FP08 cannot handle highly nonrigid skin deformations, resulting in mesh collapse, cracks or large artifacts, and tracking failures at many vertices. On the other hand, our algorithm succeeds in recovering intricate structures with dense motion information. We have performed two more comparative experiments to show the effects of the key components in the proposed algorithm. First, we have run our algorithm without the adaptive tangential rigidity term of Eq. (3), so the only regularization term is the Laplacian operator used in many other algorithms (Fig. 5). It is bit of a surprise that the system does not have a problem with the top left example in the figure, where the surface undergoes complicated nonrigid deformation, but the motion is slow and the texture information is still reliable. However, without the adaptive tangential rigidity term, the algorithm fails at recovering protruded lips where the structure and occlusions are more complex. The system also makes gross errors around eyes due to specular reflections, and on the back side of the fast moving pants, where many vertices are either not tracked or contain erroneous local motion estimates. Second, we have run our algorithm without smoothing the tangential deformation parameters (Sect. 3.2) to demonstrate the effectiveness of this smoothing step. Figure 6 shows that the algorithm without smoothing makes gross errors again at protruded lips and the back side of the pants where texture information is unreliable and local motion estimates are erroneous.

5. Conclusion and Future Work

We have presented a dense motion capture algorithm with a novel tangential rigidity constraint that models nonrigid surface deformation on tangent planes of a surface. Our experiments show that the algorithm can recover intricate surface structure and deformation such as protruded lips, facial wrinkles on the cheeks and neck, that existing algorithms cannot handle. Next on our agenda is to learn a representation of facial expressions from the reconstructed high-resolution structure and motion information, then use it to recover dense motion from new sequences acquired by one, or a few cameras. This is similar to what AAMs do, although they have been mostly used for low-resolution meshes and may not scale well or accurately capture complicated non-linear skin deformations.

Acknowledgments: This paper was supported in part by the National Science Foundation under grant IIS-0535152, the INRIA associated team Thetys, and the Agence Nationale de la Recherch under grants Hfibmr and Triangles. We thank R. White, K. Crane and D.A. Forsyth for the *pants* dataset. We also thank Hiromi Ono, Doug Epps and ImageMoversDigital for the *face* datasets.

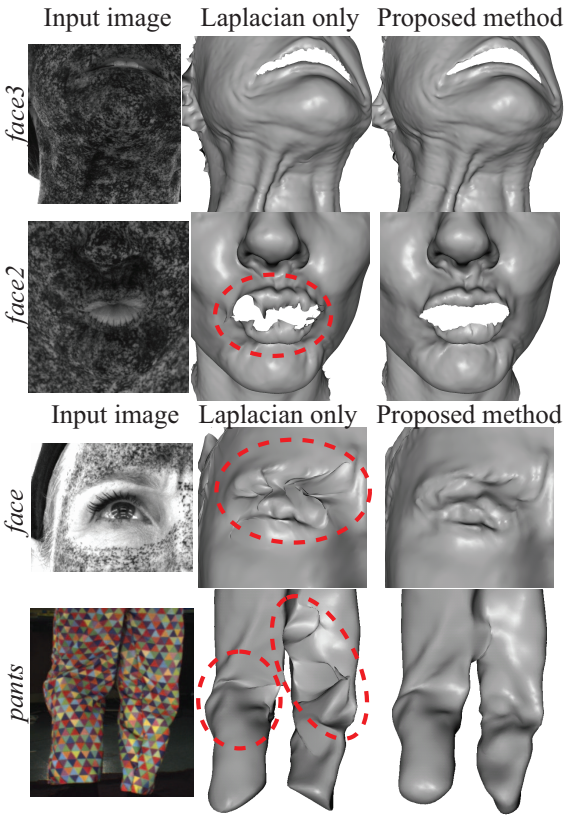

Figure 5. The *adaptive* tangential rigidity term proposed in this paper is key to filtering out erroneous local motion estimates and keeping the system stable. Without it, the algorithm does not work in three of these four examples, especially where texture information is unreliable. See text for details.

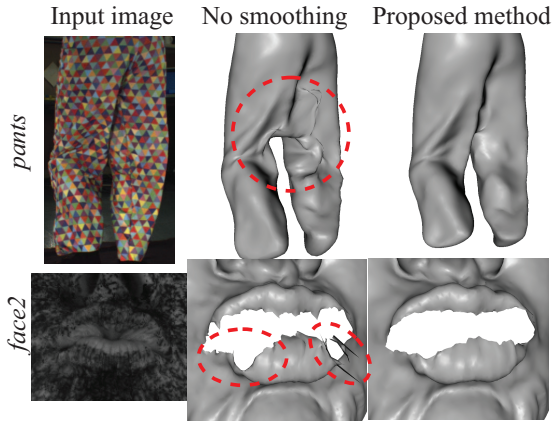

Figure 6. Smoothing tangential deformation parameters (Sect. 3.2) is essential for stability, especially at texture-poor regions.

References

[1] Dimensional imaging (http://www.di3d.com).

[2] Mova contour reality capture (http://www.mova.com).

[3] Vicon (http://www.vicon.com).

[4] B. Bickel, M. Botsch, R. Angst, W. Matusik, M. Otaduy, H. Pfister, and M. Gross. Multi-scale capture of facial geometry and motion. In *SIGGRAPH*, 2007.

[5] R. L. Carceroni and K. N. Kutulakos. Multi-view scene capture by surfel sampling: From video streams to non-rigid 3d motion, shape and reflectance. *IJCV*, 49(2-3):175–214, 2002.

[6] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. In *SIGGRAPH*, 2008.

[7] H. Delingette, M. Hebert, and K. Ikeuchi. Shape representation and image segmentation using deformable surfaces. *IVC*, 10(3):132–144, 1992.

[8] Y. Furukawa and J. Ponce. PMVS. `http://www.cs.washington.edu/homes/furukawa/research/pmvs`.

[9] Y. Furukawa and J. Ponce. Dense 3d motion capture from synchronized video streams. In *CVPR*, 2008.

[10] C. Hernández Esteban, G. Vogiatzis, G. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *ICCV*, 2007.

[11] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Symp. Geom. Proc.*, 2006.

[12] S. C. Koterba, S. Baker, I. Matthews, C. Hu, J. Xiao, J. Cohn, and T. Kanade. Multi-view aam fitting and camera calibration. In *ICCV*, volume 1, pages 511 – 518, 2005.

[13] R. Li and S. Sclaroff. Multi-scale 3d scene flow from binocular stereo sequences. In *IEEE Workshop on Motion and Video Computing*, pages 147–153, 2005.

[14] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2):135 – 164, November 2004.

[15] J. Neumann and Y. Aloimonos. Spatio-temporal stereo using multi-resolution subdivision surfaces. *Int. J. Comput. Vision*, 47(1-3):181–193, 2002.

[16] M. Odisio and G. Bailly. Shape and appearance models of talking faces for model-based tracking. In *AMFG '03*, page 143. IEEE Computer Society, 2003.

[17] S. I. Park and J. K. Hodgins. Capturing and animating skin deformation in human motion. *ACM Trans. Graph.*, 25(3):881–889, 2006.

[18] J.-P. Pons, R. Keriven, and O. Faugeras. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *IJCV*, 72(2):179–193, 2007.

[19] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. In *CVPR*, pages 2195–2202, Washington, DC, USA, 2006.

[20] K. Varanasi, A. Zaharescu, E. Boyer, and R. Horaud. Temporal surface tracking using mesh evolution. In *ECCV*, 2008.

[21] S. Vedula, S. Baker, and T. Kanade. Image-based spatio-temporal modeling and view interpolation of dynamic events. *ACM Trans. Graph.*, 24(2):240–261, 2005.

[22] D. Vlasic, I. Baran, W. Matusik, and J. Popović. Articulated mesh animation from multi-view silhouettes. In *SIGGRAPH*, 2008.

[23] R. White, K. Crane, and D. Forsyth. Capturing and animating occluded cloth. In *SIGGRAPH*, 2007.

[24] 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.