3D Structure Refinement of Nonrigid Surfaces through Efficient Image Alignment

Yinqiang Zheng, Shigeki Sugimoto, and Masatoshi Okutomi

Department of Mechanical and Control Engineering, Tokyo Institute of Technology

Abstract. Given a template image with known 3D structure, we show how to refine the rough reconstruction of nonrigid surfaces from existing feature-based methods through efficient direct image alignment. Under the mild assumption that the barycentric coordinates of each 3D point on the surface keep constant, we prove that the template and the input image are correlated by piecewise homography, based on which a direct Lucas-Kanade image alignment method is proposed to iteratively recover an inextensible surface even with poor texture and sharp creases. To accelerate the direct Lucas-Kanade method, an equivalent but much more efficient method is proposed as well, in which the most time-consuming part of the Hessian can be pre-computed as a result of combining additive and inverse compositional expressions. Sufficient experiments on both synthetic and real images demonstrate the accuracy and efficiency of our proposed methods.

1 Introduction

3D recovery of non-rigid surfaces from individual images is still a challenging task in computer vision due to its inherent ambiguity, which requires taking full advantage of available image information and other proper constraints to disambiguate the reconstruction. Such additional constraints range from physical knowledge in physics-based 3D recovery, e.g. [1, 2] among many others, to temporal consistency in 3D tracking [3, 4] and template-free recovery [5], and to geometric constraints in non-rigid 3D detection [6, 7, 8]. In this paper, we consider inextensible non-rigid surfaces and incorporate the constraints on the surface mesh edges as in [6, 8]. Our concentration is on the usage of the surface texture so as to handle sparsely textured nonrigid surfaces with sharp shape details, such as creases and folds as shown in Fig.1. Many other image cues, like silhouettes and contours ([1, 9, 10] to cite a few), have also been used for non-rigid 3D recovery, but we do not consider them here.

According to how to make use of surface texture, the majority of existing methods for non-rigid 3D recovery can be roughly categorized into two groups:

Feature-based methods: The feature-based methods establish 3D-2D feature correspondences in template-based recovery [3,4,6,7,8], or 2D-2D ones for a long video sequence in Deformable Structure from Motion [11, 12] or simply for two consecutive frames [5], and then recover 3D structure by minimizing, explicitly or implicitly, certain measurement of reprojection error. The objective function is

R. Kimmel, R. Klette, and A. Sugimoto (Eds.): ACCV 2010, Part IV, LNCS 6495, pp. 76–89, 2011.

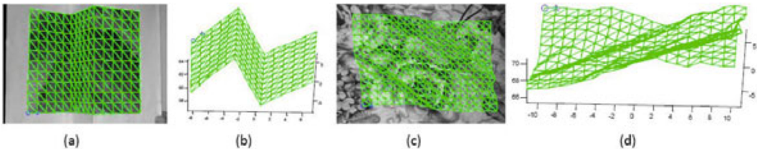

Fig. 1. Reconstruction of inextensible surfaces from single images. Existing feature-based methods tend to oversmooth the sparsely textured and sharply creased surface in (a); For a relatively well-textured surface with sharp folds (c), our image alignment method can improve its photo-consistency. (b) and (d) are their corresponding 3D meshes, respectively.

relatively easy to optimize, thus best suited to detect shape from a single image. However, to obtain a reliable reconstruction, the surface should be well-textured with dense salient features over the whole surface, which is generally not the case in practice. To tackle poorly textured surfaces, some prior cues, like the smoothness assumption [13] and the deformation model [6, 7, 8], are frequently introduced, whose results are roughly correct but weak in photo-consistency, like Fig.1(c). More seriously, such prior knowledge tends to oversmooth sharp details, thus can not be used to accurately recover sharply creased surfaces. Another possible alternative is to establish dense correspondences through non-rigid 2D registration as in [3, 6]. However, even the state-of-the-art methods [14, 15] introduce some shape terms to penalize sharp deformation, thereby inapplicable for sharp creased and folded surfaces.

Appearance-based methods: The appearance-based methods (or direct methods) take advantage of the intensity of each pixel of the surface image, and consequently are able to obtain photo-consistent reconstruction even for poorly textured surfaces. Unfortunately, the resulting objective function is highly non-convex, thereby commonly used in a tracking scenario. Another well-known drawback of the appearance-based methods is its inefficiency due to re-evaluation of the Hessian for each pixel at each iteration. Interestingly, some efficient Inverse Compositional Image Alignment (ICIA) algorithms have been proposed for 3D tracking of human face [16]. To the best knowledge of the authors, no such efficient algorithm exists for generic non-rigid surfaces.

In order to accurately reconstruct (or more exactly detect) the 3D structure of sparsely textured surfaces from a single image, it is desirable to fuse the featured-based and appearance-based methods, so that we can initialize the non-convex image alignment by using the easy-to-solve feature-based methods and derive photo-consistent shape from each pixel of the surface image, rather than from the prior knowledge. Such fusion has been proved feasible for fast non-rigid 2D recovery [14]. In this paper, we extend it to 3D case.

Given a template image whose 3D structure is known, we follow the feature-based robust convex method combined with local deformation model [8] and the closed-form solution with global deformation model [6] to derive rough reconstructions, which are used to initialize the iterative appearance-based methods.

Under the mild assumption that the barycentric coordinates of each point on its corresponding patch of the 3D surface are constant, a basic assumption underlying many non-rigid 3D surface recovery methods [3,4,6,7,8], we show that the template image and the input image are correlated by piecewise homograpy. Based on this homography warp, we propose a direct Lucas-Kanade method, also known as Forward Additive Image Alignment method [17], to recover sparsely textured surfaces by integrating the constraints on mesh edges to disambiguate the reconstruction. Since we never introduce the smooth term or the deformation model at this step, the direct image alignment method enables us to tackle sharp details with strong photo-consistency. The well-known downside of the direct Lucas-Kanade image alignment lies in its inefficiency, since the Jacobian and Hessian should be recomputed for each pixel at each iteration. To improve efficiency, an equivalent method is proposed as well by combing the additive and inverse compositional expressions. Although the Hessian is not completely constant across iteration, the most time-consuming part can be computed offline, while the inconstant part needs only to be recomputed for each patch of the surface, not for each pixel on the input image. This makes it much faster than the direct Lucas-Kanade method, although the folds of acceleration are dependent on the resolution of the surface mesh. In a typical experiment with a 11x16 triangulated mesh and 720x576 images, it takes about 0.2s to compute the Hessian, in contrast to 2.8s in the direct Lucas-Kanade method.

In the remaining of the paper, we first derive the warp function between the template and the input image in section 2, based on which the appearance-based methods for non-rigid 3D recovery are introduced in section 3. We present the fusion method in Section 4. Section 5 shows the extensive experimental results for both synthetic and real images and section 6 includes some concluding remarks of this paper.

2 Warp between the Template and the Input Image

In this section, we disclose the warp function that relates the template and the input image. Before that, we introduce the notations and assumptions we made.

2.1 Notations and Assumptions

The deformable surface is explicitly parameterized as a triangulated mesh, as in Fig.1, with N_v vertices $V_i = (x_i, y_i, z_i)^T$, $1 \leq i \leq N_v$. The unknown to be estimated is X, a column vector obtained by concatenating x-,y-,z- coordinates of the N_v vertices. Specifically, $X = \left[x_1\ y_1\ z_1 \cdots x_{N_v}\ y_{N_v}\ z_{N_v}\right]^T$. The known 3D mesh corresponding to the template image is denoted by $\tilde{X}$, thereafter named as template 3D mesh for short. The mesh is composed of N_p patches and N_e edges. For patch $j, 1 \leq j \leq N_p$, its three vertices are noted by V_{j1}, V_{j2} and V_{j3}, whose corresponding vertices in the template 3D mesh are $\tilde{V}_{j1}$, $\tilde{V}_{j2}$ and $\tilde{V}_{j3}$, respectively. For simplicity, all these vertices are in the camera referential without loss of generality.

The mesh is assumed to be flexible but inextensible, thus preventing the distance between two neighboring vertices from expanding or shrinking too much. We also assume a pinhole perspective camera model, whose intrinsic parameter matrix K is known and keep constant.

2.2 2D-2D Warp Function

The 3D warp of a non-rigid mesh is often modeled as piecewise affine transformation. Specifically, it is usually assumed, as in [3,4,8,6,7], that the barycentric coordinates of each point on its corresponding 3D patch are constant across deformation. The constancy of barycentric coordinates indicates that the surface patches are always planar across deformation, thus one can easily expect that the 2D warp between the template and the input image is piecewise homography. In the following, we present the explicit formulation of the 2D warp.

For any point $\tilde{Q}$ on the jth patch of the template 3D mesh, its coordinates can be expressed as

$$\tilde{Q} = \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right] \varepsilon, \tag{1}$$

where ε is the barycentric coordinates of $\tilde{Q}$ on this patch. Its homogeneous projection $\tilde{q}$ on the template image T is:

$$\tilde{\lambda}\tilde{q} = K \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right] \varepsilon, \tag{2}$$

with unknown scalar $\tilde{\lambda}$ accounting for the depth.

On the unknown 3D mesh corresponding to the input image I, $\tilde{Q}$ is transferred to Q after some deformation. Since we assume its barycentric coordinates keep constant, the coordinates of Q can be written as

$$Q = \left[V_{j1}\ V_{j2}\ V_{j3}\right] \varepsilon, \tag{3}$$

whose homogeneous projection q on the input image I should be:

$$\lambda q = K \left[V_{j1}\ V_{j2}\ V_{j3}\right] \varepsilon, \tag{4}$$

where λ is a unknown depth scalar.

Combing eq.(1)-(4), we get

$$(\lambda/\tilde{\lambda})q = K \left[V_{j1}\ V_{j2}\ V_{j3}\right] \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right]^{-1} K^{-1}\tilde{q}. \tag{5}$$

Considering that V_{jk} and $\tilde{V}_{jk}, 1 \leq k \leq 3$ are the vertices of the jth triangular patch, the 3x3 matrix P is invertible, where

$$P = K \left[V_{j1}\ V_{j2}\ V_{j3}\right] \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right]^{-1} K^{-1}, \tag{6}$$

meaning that the 2D-2D warp P is a homography. Therefore, the template image T and the input image I are correlated by piece-wise homography.

3 Appearance-Based Non-rigid 3D Recovery

Based on the 2D warp, we show how to directly use appearance-based image alignment to recover sparsely textured surface from a single image.

3.1 Direct Lucas-Kanade Method for Non-rigid 3D Recovery

The direct Lucas-Kanade image alignment method [18] is an iterative method to minimize the Sum of Squared Difference (SSD) between the template image T and the input image I by additively adjusting the parameters from a given starting point.

Minimizing SSD. The direct Lucas-Kanade method uses an additive update for the unknown parameter $X \leftarrow X + \Delta X$, where ΔX is the increment in current iteration. The Warp W for the jth patch is P, which can be updated as:

$$P = K \left[V_{j1} + \Delta V_{j1} \; V_{j2} + \Delta V_{j2} \; V_{j3} + \Delta V_{j3} \right] \left[\tilde{V}_{j1} \; \tilde{V}_{j2} \; \tilde{V}_{j3} \right]^{-1} K^{-1}. \tag{7}$$

Under the assumption of constant intensity, the increment ΔX can be solved by minimizing the SSD energy term $E_{SSD}(X)$:

$$min_{\Delta X} E_{SSD}(X + \Delta X) = \sum_{j=1}^{N_p} \sum_{u \in C_j} \left[I(W(u, X + \Delta X)) - T(u) \right]^2, \tag{8}$$

where C_j is the set of pixels in the image of the jth patch. After using Gauss-Newton approximation, the increment ΔX can be calculated by

$$\Delta X = H_{SSD}^{-1} \{ \sum_{j=1}^{N_p} \sum_{u \in C_j} [\nabla I \frac{\partial W}{\partial \Delta X}]^T [T(u) - I(W(u; X))] \}, \tag{9}$$

where the Hessian H_{SSD} for the SSD energy term $E_{SSD}(X)$ should be

$$H_{SSD} = \sum_{j=1}^{N_p} \sum_{u \in C_j} [\nabla I \frac{\partial W}{\partial \Delta X}]^T [\nabla I \frac{\partial W}{\partial \Delta X}]. \tag{10}$$

Ambiguity Analysis. According to eq.(6), there are 9 unknowns in the homography, whereas the homography has only 8 independent parameters. Therefore, even assuming perfect alignment for each pixel in the projection of this patch, there is one scalar ambiguity in the estimated patch coordinates. For the whole triangulated mesh, assuming perfect alignment for each patch and considering the connectivity between mesh patches, ideally there is still a global scalar ambiguity in the reconstruction if we only minimize the SSD energy term E_{SSD} as in eq.(8). Actually, we find in our experiment that the Hessian H_{SSD} is minus 1 rank-deficient with some (about one third) close-to-zero eign-values, demonstrating that monocular recovery of deformable surfaces is an ill-posed problem.

Our observation is consistent with the ambiguity analysis in [19] on the basis of dense and uniform feature correspondences. This is understandable since the correct image alignment can be regarded as establishing extremely dense correspondences, i.e. one feature correspondence for one pixel. Although the image alignment does not better constrain the reconstruction for a well-textured surface, we can indeed expect that it works better for sparsely textured surfaces, the recovery of which becomes more under-constrained when using sparse correspondences only.

Disambiguating Reconstruction. To obtain an unique and stable reconstruction, we introduce constraints on each edge of the mesh by penalizing it from expanding and shrinking too much. For the k th, $1 \leq k \leq N_e$, edge of the mesh defined by two neighboring vertices V_{k1} and V_{k2}, the constraints can be written as: $||V_{k1} - V_{k2}|| = l_k$, where $|| \cdot ||$ represents L_2 norm, and l_k is the length of kth edge in the template 3D mesh. It can be rearranged into matrix form $||S_k X|| = l_k$. Rather than using them as hard constraints, we minimize the equivalent side-length energy term $E_s(X)$, which is defined by

$$E_s(X) = \sum_{k=1}^{N_e} (||S_k X||^2 - l_k^2)^2. \tag{11}$$

Combined with the SSD energy term $E_{SSD}(X)$, the direct Lucas-Kanade image alignment method can be formulated as:

$$min_{\Delta X}\{E_{SSD}(X + \Delta X) + \omega_s E_s(X + \Delta X)\}, \tag{12}$$

where ω_s is a user-defined weighting factor. Using Gauss-Newton approximation, the increment ΔX can be easily calculated.

Without introducing any *a priori* knowledge that tend to oversmooth sharp details, our appearance-based method can be used to accurately recover inextensible surfaces with poor texture and sharp creases.

From eq.(10), we can see that the Jacobian and the Hessian should be recomputed for each pixel at each iteration, since they are evaluated at current estimation of the vertex parameters X. Generally it is computationally demanding. In the following, we show how to accelerate this direct Lucas-Kanade method by using ICIA, in which the most time-consuming part can be precomputed.

3.2 Efficient Image Alignment for Non-rigid 3D Recovery

Combining Additive and Inverse Compositional Expressions. In ICIA [17], the warp is updated by $W \leftarrow \bar{W} \circ (\Delta W)^{-1}$, where the operator '$\circ$' means the composition of the current warp $\bar{W}$ and the increment warp ΔW. Specifically, for a homography warp, it can be updated by $P \leftarrow \bar{P}(I + \Delta P)^{-1}$, where $\Delta W = I + \Delta P$ is the incremental homography warp, and $\bar{W} = \bar{P}$ is the current homography.

Since we need to estimate the vertex coordinates embedded in the homography, rather than the homography in itself, we have to devise the update rule for the mesh parameters X. Same as the direct Lucas-Kanade method, we use an additive update rule for X, i.e. $X \leftarrow X + \Delta X$. To make the warp updated from the inverse composition equivalent to that from the additive updating in eq.(7), we let the following equation hold:

$$P = \bar{P}(I + \Delta P)^{-1}, \tag{13}$$

where P is from eq.(7), while $\bar{P}$ from eq.(6). This rule has been used in [20] for fast surface reconstruction from stereo. Note that it is not completely the same as the original ICIA image alignment in [17], since the parameter X can be directly updated through the additive rule. In the following, we still name our method as an ICIA method, considering that the homography warp is updated by inverse composition.

It is obvious that when $\Delta X \to 0$, the incremental warp $\Delta W \to I$, which means that it is an identity warp. Before giving the explicit relationship between ΔP and ΔX, we first show how to use the ICIA method in non-rigid 3D recovery.

According to [17], image alignment can alternatively be formulated as:

$$min_{\Delta X} E_{SSD}(X + \Delta X) = \sum_{j=1}^{N_p} \sum_{u \in C_j} \left[T(\Delta W(u, \Delta X)) - I(W(u, X))\right]^2. \tag{14}$$

Using Gauss-Newton approximation, the increment ΔX can be derived from:

$$\Delta X = H_{SSD}^{-1}\{\sum_{j=1}^{N_p} \sum_{u \in C_j} [\nabla T \frac{\partial \Delta W}{\partial \Delta p} \frac{\partial \Delta p}{\partial \Delta X}]^T [T(u) - I(W(u; X))]\}, \tag{15}$$

where Δp represents the elements in ΔP, and the Hessian

$$H_{SSD} = \sum_{j=1}^{N_p} \sum_{u \in C_j} [\nabla T \frac{\partial \Delta W}{\partial \Delta p} \frac{\partial \Delta p}{\partial \Delta X}]^T [\nabla T \frac{\partial \Delta W}{\partial \Delta p} \frac{\partial \Delta p}{\partial \Delta X}]. \tag{16}$$

To calculate H_{SSD}, we need to compute $\partial \Delta p / \Delta X$, which is presented in the following subsection.

Computing $\boldsymbol{\partial \Delta p / \Delta X}$. When $\Delta P \to 0$, the inverse of the incremental warp can be approximated (first order approximation) by

$$(I + \Delta P)^{-1} = I - \Delta P. \tag{17}$$

From eq.(13) and eq.(17), we can calculate ΔP as follows:

$$\Delta P = -K \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right] \left[V_{j1}\ V_{j2}\ V_{j3}\right]^{-1} \left[\Delta V_{j1}\ \Delta V_{j2}\ \Delta V_{j3}\right] \left[\tilde{V}_{j1}\ \tilde{V}_{j2}\ \tilde{V}_{j3}\right]^{-1} K^{-1}, \tag{18}$$

from which, the $\partial \Delta p / \Delta X$ can be straightforwardly computed, since Δp is a linear function of ΔX.

Efficiency Analysis. From eq.(16), the gradient of the template image ∇T and that of the increment warp $\partial \Delta W / \Delta p$, i.e. the most time-consuming pixel-related parts of the Hessian, are constant across iteration, since they are evaluated at $\Delta X = 0$. However, $\partial \Delta p / \Delta X$ is dependent on the current estimation of X, thus should be recomputed at each iteration. Fortunately, it is irrelevant to pixel coordinates, and needs only to be recomputed for each patch. Specifically,

$$H_{SSD} = \sum_{j=1}^{N_p} (\frac{\partial \Delta p}{\partial \Delta X})^T H_{const} \frac{\partial \Delta p}{\partial \Delta X}, \tag{19}$$

where H_{const} is the constant part of the Hessian,

$$H_{const} = \sum_{u \in C_j} [\nabla T \frac{\partial \Delta W}{\partial \Delta p}]^T [\nabla T \frac{\partial \Delta W}{\partial \Delta p}]. \tag{20}$$

To disambiguate the reconstruction, we should introduce the side-length energy term $E_s(X)$ as in section 3.1. The Hessian for this term should also be recomputed. However, the number of sides is always much smaller than that of pixels, thus can be evaluated very fast.

4 Fusing Features and Appearance

The appearance-based image alignment methods are usually sensitive to disturbance on the pixel intensity. When lighting changes or small occlusion occurs, it is helpful to fuse feature correspondences and appearance-based image alignment [14], since these feature points, serving somewhat as anchors, are able to prevent the mesh from drifting. In addition, introducing feature correspondences poses little increase in computational burden, since the Hessian for this part can be easily computed. The feature set used here is the inlier set from the feature-based methods whose reprojection error is lower than 1 pixel. The feature energy term $E_f(X)$ is measured by

$$E_f(X) = ||MX||^2, \tag{21}$$

where M is the structure matrix constructed by following [6,7,8]. Specifically, we simultaneously minimize the SSD energy term $E_{SSD}(X)$, the side-length term $E_s(X)$ and the feature energy term $E_f(X)$:

$$min_{\Delta X}\{E_{SSD}(X + \Delta X) + \omega_s E_s(X + \Delta X) + \omega_f E_f(X + \Delta X)\}, \tag{22}$$

where ω_f is a user-defined weighting factor. This equation can be easily solved by using Gauss-Newton approximation. Note that the Hessian for the SSD energy term E_{SSD} can be calculated either by eq.(10) in the direct Lucas-Kanade method or by eq.(19) in the ICIA method.

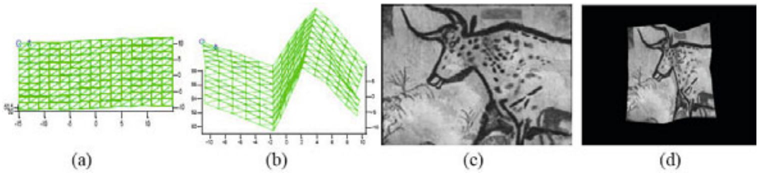
(a) (b) (c) (d)

Fig. 2. Synthetic 100-frame mesh sequence. (a) The mesh in rest position. (b) The 50th frame with largest deformation. The sequence is retextured by using sparse texture (c), and reprojected onto images (d) by a virtual camera.

5 Experimental Results

In this section, we use both synthetic data and real images to test the performance of our proposed methods. For feature-based methods, we use 60 modes for the global deformation model [6] and 20 modes for each of the local deformation model [8].

5.1 Synthetic Data

We generate a 100-frame sequence of a piece of paper with sharp creased deformation by using motion capture devices (Fig.2(a,b)), then we synthesize sparse texture (Fig.2(c)) on the meshes, which are backprojected onto 2D images using a synthetic projective camera (Fig.2(d)). The mesh resolution is 11x16, and size is 200mm x 300mm.

Convergence w.r.t. Rough Initialization and Intensity Noise. Here, we use the 50th frame as the target, whose deformation is the largest in the sequence. We add zero mean Gaussian noise with deviation σ on the ground-truth 3D mesh to simulate rough initialization. Both the template and the input image are corrupted by zero mean Gaussian noise with deviation 2 grey levels. 100 sparse feature correspondences are also randomly generated for the fusion methods. We measure the average vertex-to-vertex error between the ground-truth 3D mesh and the estimated 3D mesh from image alignment after 20 iterations. The result is said to be convergent when the average 3D error is lower than 2mm. We compare the performance of the direct Lucas-Kanade method (DLK), the Inverse Compositional Image Alignment method (ICIA), and their fusion with features, shown in Fig.3(a) as (F+DLK) and (F+ICIA), respectively. We vary σ from 0.4 mm to 4 mm, and repeat each method for 500 times at each noise level. From Fig.3(a), we can see that the DLK and the ICIA have almost the same performance, which is understandable since they are almost equivalent. When noise is large, the ICIA method is slightly weaker than the DLK method due to the first-order approximation used in eq.(17). Both methods can be improved by fusing feature correspondences.

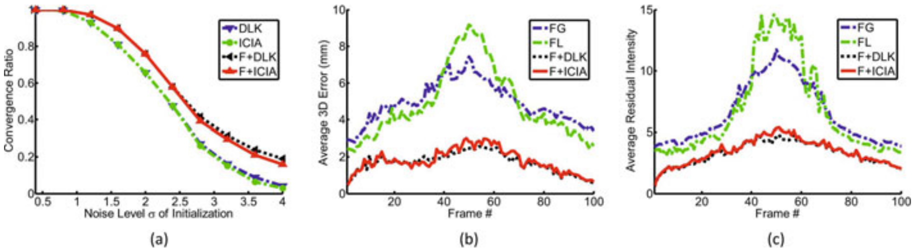

Fig. 3. Synthetic experiments. (a) Use the 50th frame (with largest deformation) to test the convergence performance w.r.t. rough initialization with intensity noise. (b) The average 3D vertex-to-vertex error for the whole 100-frame sequence from four different methods. (c) The average intensity of the residual image for four different methods. We can see that when large deformation occurs (in the middle of the sequence), the feature-based methods (FG) and (FL) can not accurately recover the mesh, which can be improved by our fusion methods (F+DLK) and (F+ICIA).

Table 1. Time Performance in direct Lucas-Kanade (DLK) and ICIA methods

Methods	Precomputation	Compute Hessian	One Iteration
DLK	-	2.827s	3.462s
ICIA	2.672s	0.212s	0.718s

In the following, we initialize the appearance-based methods by using the rough results from feature-based methods. Considering that the fusion methods work better, we shall only use the two fusion ones (F+DLK) and (F+ICIA).

Improving Feature-Based Methods by Fusion. Here we use SIFT [21] to establish 3D-2D feature correspondences for the whole sequence, and follow the feature-based closed-form solution with global deformation model (FG) [6] and the convex method with local deformation model (FL) [8] to get rough initialization. The fusion methods (F+DLK) and (F+ICIA) are initialized by both feature-based methods (FG) and (FL), and only the results with less residual intensity are presented. Both the template and the input image are corrupted by Gaussian noise with zero mean and deviation 2 gray levels. Fig.3(b) shows the average 3D vertex-to-vertex error and Fig.3(c) the average intensity on the residual image. The results of the 50th frame are also presented in Fig.4. From these results, we can see that, compared with the local deformation model, the global deformation model is more likely to oversmooth the sharp creases. However, when large deformation occurs, the shape from the local deformation model is somewhat irregular. This is due to the fact that the local deformation model relies not so heavily on the training data. These problems can be reliably remedied by our fusion methods, both of which have almost the same performance.

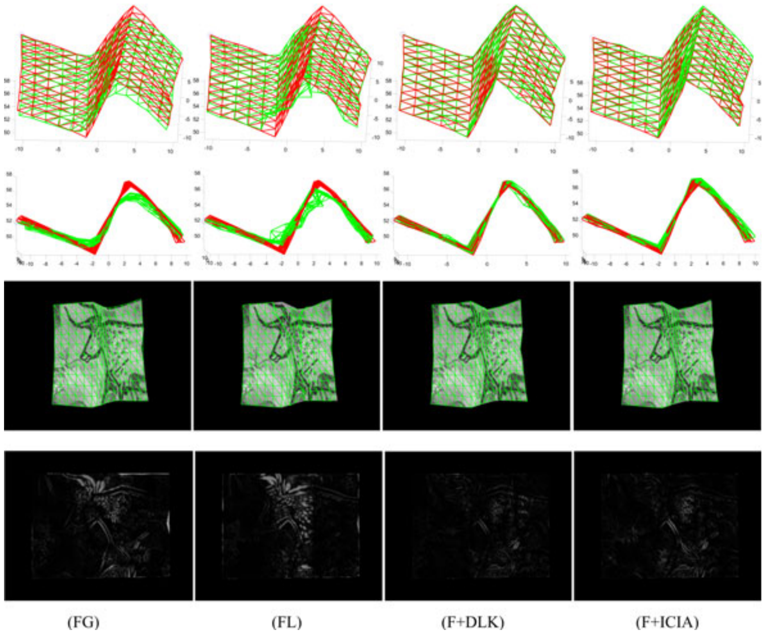

Fig. 4. Visual comparison of the results from different methods for the 50th frame with largest deformation. From left to right are the results from the global deformation model (FG), the local deformation model (FL), the fusion method with direct Lucas-Kanade (F+DLK), and the fusion one with ICIA (F+ICIA), respectively. From top to bottom are the recovered 3D mesh(green) together with the ground truth(red), the same 3D meshes observed from side view, the input image overlaid by the mesh projection, and the residual images, respectively.

Time Performance. We implement all the methods in a 1.6GHz laptop with 2GB RAM by using MATLAB. The image resolution is 720x576, and the mesh size is 11x16 with 528 variables, 475 edges and 300 patches. The time performance is shown in Table 1 for one iteration averaged from 50 iterations. The precomputation of the constant part of the Hessian for the ICIA method takes 2.672s. It only takes 0.212s in the ICIA method to compute the full Hessian, in contrast to 2.827s in the direct Lucas-Kanade method, which is about 14 times faster. The ICIA method takes 0.718s in one iteration, while the direct Lucas-Kanade method takes 3.462s, with an acceleration rate about 5 times. The acceleration rate shrinks, because some other pixel-related process, like bilinear interpolation, takes about 0.4s, which is a bottleneck when using MATLAB.

5.2 Real Images

We also have some preliminary results on a piece of paper with sparse texture and sharply creased deformation. To show that our methods can accurately

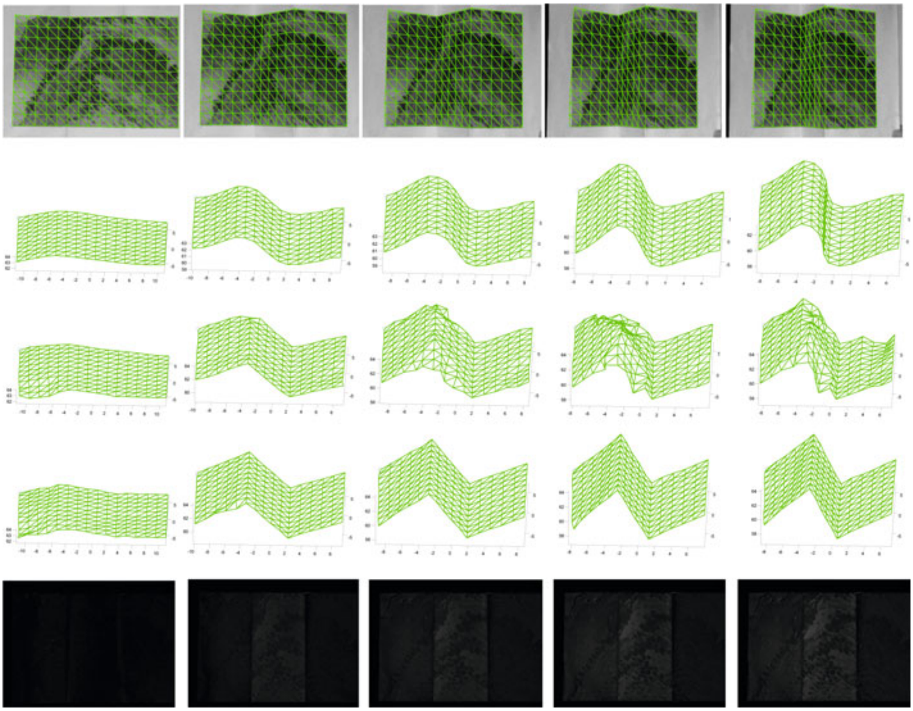

Fig. 5. Five frames of a piece of paper with sparse texture and sharp creases. First row: Images overlaid by the projection of the mesh reconstructed from our method. From 2nd to 4th row: 3D meshes from the Global deformation model (FG), Local deformation model (FL), and our fusion method (F+ICIA), respectively. Last row: The residual images after image alignment.

recover the sharp creases, we intentionally make the creases coincident with the mesh edges, otherwise the recovered creases would be smoothed due to surface discretization. The images are captured by a FLea2 camera with 800x600 resolution. Generally the matched 3D-2D feature pairs are less than 200. Considering the efficiency of the (F+ICIA) method and its equivalence to the (F+DLK) method, we only use the (F+ICIA) method here, which is initialized by the (FL) method. From Fig.5, we can see again that the global deformation model can only approximate the sharply creased surface, while the shape from the local deformation model is somewhat irregular due to severe lack of features, although not so seriously being oversmoothed. By observing the residual images in the last row, we see that our fusion method works well in case of local intensity changes caused by significant variation in surface orientation. We also show our method (F+ICIA) can improve the photo-consistency of the results from the (FL) method [8] for a piece of cloth with relatively dense texture and sharp folds, which can be concluded by comparing the residual images in Fig.6.

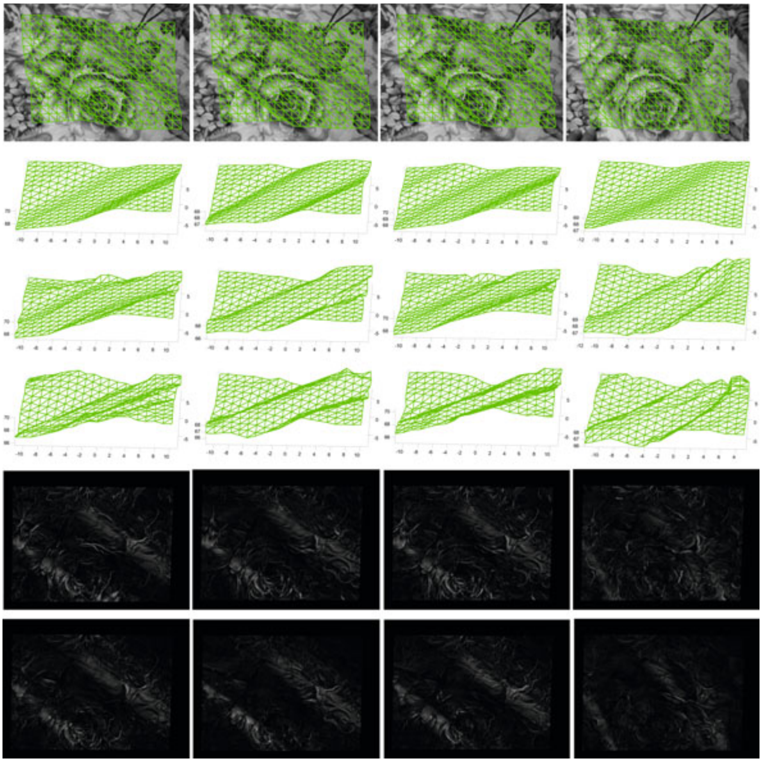

Fig. 6. Four frames of a piece of cloth with relatively dense texture and sharply folded deformation. First row: Images overlaid by the projection of the mesh reconstructed from our method (F+ICIA). From 2nd to 4th row: 3D meshes from Global deformation model (FG), Local deformation model (FL), and our method (F+ICIA), respectively. From 5th to 6th row: The residual images for (FL) and for (F+ICIA), respectively.

6 Conclusion

We have shown how to efficiently refine the 3D structure of poorly textured nonrigid surfaces, even with sharp details, from a single image by fusing feature correspondences and appearance-based image alignment. To our knowledge, this work is the first one that can accurately recover sharply creased surfaces in case of sparse texture.

We should mention that it is still a challenging task to recover sharp details in case of large occlusion. When large occlusion occurs, it is inevitalbe to introduce some prior knowledge, which tends to oversmooth sharp details. In addition, we partially conquer the disturbance on pixel intensity by fusing features, which is insufficient in case of large lighting changes. The potential lighting variation can be further compensated, by using the Dual ICIA method [22] for efficiency, which is left to the future.

Acknowledgement. This work was partly supported by Grant-in-Aid for Scientific Research (21240015) from the Japan Society for the Promotion of Science.

References

1. Cohen, L., Cohen, I.: Finite element methods for active contour models and balloons for 2d and 3d images. PAMI 15, 1131–1147 (1993)
2. Bhat, K.S., Twigg, C.D., Hodgins, J.K., Khosla, P.K., Popovic, Z., Seitz, S.M.: Estimating cloth simulation parameters from video. In: ACM Symposium on Computer Animation (2003)
3. Salzmann, M., Hartley, R., Fua, P.: Convex optimization for deformable surface 3d tracking. In: ICCV (2007)
4. Zhu, J., Hoi, S.C.H., Xu, Z., Lyu, M.R.: An effective approach to 3d deformable surface tracking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 766–779. Springer, Heidelberg (2008)
5. Varol, A., Salzmann, M., Tola, E., Fua, P.: Template-free monocular reconstruction of deformable surfaces. In: ICCV (2009)
6. Salzmann, M., Moreno-Noguer, F., Lepetit, V., Fua, P.: Closed-form solution to non-rigid 3d surface registration. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 581–594. Springer, Heidelberg (2008)
7. Moreno-Noguer, F., Salzmann, M., Lepetit, V., Fua, P.: Capturing 3d stretchable surfaces from single images in closed form. In: CVPR (2009)
8. Salzmann, M., Fua, P.: Reconstructing sharply folding surfaces: a convex formulation. In: CVPR (2009)
9. Ilic, S., Salzmann, M., Fua, P.: Implicit meshes for effective silhouette handling. IJCV 72, 159–178 (2007)
10. Terzopoulos, D., Metaxas, D.: Dynamic 3d models with local and global deformations: deformable superquadrics. PAMI 13, 703–714 (1991)
11. Vidal, R., Hartley, R.: Perspective nonrigid shape and motion recovery. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 276–289. Springer, Heidelberg (2008)
12. Xiao, J., Kanade, T.: Uncalibrated perspective reconstruction of deformable structures. In: ICCV (2005)
13. Ecker, A., Jepson, A., Kutulakos, K.: Semidefinite programming heuristics for surface reconstruction ambiguities. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 127–140. Springer, Heidelberg (2008)
14. Zhu, J., Lyu, M.R., Huang, T.S.: A fast 2d shape recovery approach by fusing features and appearance. PAMI 31, 1210–1224 (2009)
15. Pilet, J., Lepetit, V., Fua, P.: Fast non-rigid surface detection, registration and realistic augmentation. IJCV 76, 109–122 (2008)
16. Munoz, E., Buenaposada, J., Baumela, L.: Efficient model-based 3d tracking of deformable objects. In: ICCV (2005)
17. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. IJCV 56, 221–255 (2004)
18. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. IJCAI (1981)
19. Salzmann, M., Lepetit, V., Fua, P.: Deformable surface tracking ambiguities. In: CVPR (2007)
20. Sugimoto, S., Okutomi, M.: A direct and efficient method for piecewise-planar surface reconstruction. In: CVPR (2007)
21. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 20, 91–110 (2004)
22. Bartoli, A.: Groupwise geometric and photometric direct image registration. PAMI 30, 2098–2108 (2008)