Non-Rigid Shape-from-Motion for Isometric Surfaces using Infinitesimal Planarity

Ajad Chhatkuli

Daniel Pizarro

Adrien Bartoli
http://isit.u-clermont1.fr/~ab/

ALCoV-ISIT, UMR 6284 CNRS / Université d'Auvergne, Clermont-Ferrand, France

Abstract

This paper proposes a general framework to solve Non-Rigid Shape-from-Motion (NRSfM) with the perspective camera under isometric deformations. Contrary to the usual low-rank linear shape basis, isometry allows us to recover complex shape deformations from a sparse set of images. Existing methods suffer from ambiguities and may be very expensive to solve. We bring four main contributions. First, we formulate isometric NRSfM as a system of first-order Partial Differential Equations (PDE) involving the shape's depth and normal field and an unknown template. Second, we show this system cannot be locally resolved. Third, we introduce the concept of infinitesimal planarity and show that it makes the system locally solvable for at least three views. Fourth, we derive an analytic solution which involves convex, linear least-squares optimization only, and outperforms existing works.

1 Introduction

The 3D reconstruction of rigid scenes from multiple images has been extensively studied over the last few decades. Rigidity strongly constrains the geometry of multiple images and lies at the heart of Shape-from-Motion (SfM) [9]. However, rigid reconstruction techniques fail when applied directly to deforming objects such as the human body or a piece of cloth. Shape-from-Template (SfT) reconstructs shape from a single image given a 3D shape template [2, 15] and thus handles deforming shapes. Non-Rigid Shape-from-Motion (NRSfM) is the general problem of reconstructing a deforming 3D shape from multiple monocular images [16, 19]. NRSfM is a much harder problem than SfM and SfT and still involves unresolved challenges. Both SfT and NRSfM are based on prior object constraints from two categories *i)* shape space constraints, usually modeled by statistical models such as the low-rank shape model [6, 17] and *ii)* deformation space constraints, usually derived from physics such as isometry [1, 15], conformity [2] and linear elasticity [12]. Existing SfT methods have been extensively based on isometry, which has been shown to be a very good model for many types of materials.

Using isometry in NRSfM is a very appealing idea. However, most of the recent solution attempts suffer from theoretical or practical problems. [5, 16] use the orthographic camera

to recover the shape's normal locally; they suffer from local two-fold ambiguities and significantly degrade for shorter focal lengths. [18] specifically addresses the case of piecewise planar surfaces; it uses the perspective camera but still has patch-wise two-fold unresolved ambiguities induced by the processing of image pairs. [19] solves NRSfM for both perspective and orthographic cameras with a relaxed version of the isometry constraint but uses a costly optimization process.

We propose a general framework to formulate and solve NRSfM with the perspective camera and under isometric deformations. Our contributions are fourfold. First we give a new formulation of isometric NRSfM as a nonlinear system of first-order Partial Differential Equations (PDE) involving the shapes' normal and depth functions. Our formulation includes an unknown template, and has SfT as a special case. Second, we show that independent solutions for depth and normal in our system of PDEs are underconstrained. This is an important result since it tells us that NRSfM cannot be solved locally using only first-order PDEs, in contrast to SfT [2]. Third, we introduce infinitesimal planarity which assumes that the shape's second-order derivatives are zero pointwise. We show that this makes the system of PDEs locally solvable for at least three views. Fourth, we propose an algorithm which involves only linear least-squares. We show that our method outperforms existing work on challenging synthetic and real data with groundtruth.

2 Previous Work

Current NRSfM methods may be grouped into two categories. The first category uses statistical shape space priors. They usually model the shape space with the low-rank shape model [3]. It may be learnt a priori [7] or jointly reconstructed with shape and motion, as in the so-called non-rigid factorization methods [6, 17]. These methods usually use supplemental priors such as temporal smoothness [17]. A very related framework is the one of trajectory priors [8]. Methods based on the low-rank shape model handle objects such as faces, but are not well-adapted to objects with very large shape spaces such as pieces of paper. They usually require a large number of images. The second category of methods uses physics-based deformation priors. [16] addresses the problem of isometric NRSfM for local patches with the orthographic camera. [5] shows that local reconstruction with the orthographic cameras is ambiguous and uses temporal smoothness to disambiguate. [18] uses the perspective camera to reconstruct the shape's normals for image pairs based on local homography decomposition [11]. The surfaces obtained for all image pairs are then registered to complete reconstruction. This method is promising but has two main problems. First, disambiguating the reconstructions obtained from image pairs is unresolved as smoothness turns out to be unstable in many practical cases. Second, the computation of a local homography requires one to find a large enough image support, bounding the method to handle piecewise planar deformations. [19] uses graph-based energy optimization with a discretized shape space. Although its energy function handles missing correspondences and smoothness priors the proposed energy is nonconvex, and the solution thus may not be guaranteed to be optimal.

In contrast, our solution to isometric NRSfM is analytical and involves rounds of convex, linear least-squares optimization only. We introduce a theoretical model that includes SfT as a special case and that fills most of the gaps found in previous works. We use the perspective camera model and, unlike [18] that assumes piecewise planarity, our concept of infinitesimal planarity allows us to model general smooth shapes and deformations. We show that the

solution is unique for more than two views with no two-fold ambiguities in normals and give a general reconstruction method for $n > 2$ views.

3 A General Framework for Isometric Surfaces

We first review isometric SfT as in [1, 2]. We then extend SfT to NRSfM by adding more views and keeping the template as an additional unknown. We finally analyze the existence of local solutions of the NRSfM system.

3.1 SfT: Shape-from-Template

Figure 1.a shows a general diagram for SfT whose solution is based on the reprojection and the deformation constraints [2]. The known template is represented by a 2D domain $\mathcal{T}$ corresponding to the 3D template's conformal flattening. The deformed shape $\mathcal{S}$ is modeled by an unknown embedding $\varphi \in C^2(\mathcal{T};\mathbb{R}^3)$, and $\mathcal{I}$ is an image of $\mathcal{S}$. We use Π to denote perspective projection to coordinates normalized with respect to the camera intrinsics. The registration between $\mathcal{T}$ and $\mathcal{I}$ is known and modeled by an image warp $\eta \in C^2(\mathcal{T};\mathbb{R}^2)$. The *reprojection constraint* is then $\eta = \Pi \circ \varphi$. Let $\varphi = (\varphi_x\ \varphi_y\ \varphi_z)^\top$ where $\varphi_x, \varphi_y, \varphi_z \in C^2(\mathcal{T};\mathbb{R})$ model each dimension of φ.

If $\mathcal{S}$ results of an isometric deformation of the 3D template, and since $\mathcal{T}$ was obtained by conformal flattening, the *deformation constraint* is that the first fundamental form of φ is a scaled identity matrix [2]:

$$\mathbf{J}_\varphi^\top \mathbf{J}_\varphi = \lambda^2 \mathbf{I}_{2\times 2}, \tag{1}$$

where $\mathbf{J}$ is the first-order partial derivatives operator and $\lambda \in C^2(\mathcal{T};\mathbb{R}^+)$ is the flattening scale. As the two columns of $\mathbf{J}_\varphi$ are orthogonal we may rewrite Eq. (1) as:

$$(\mathbf{J}_\varphi\ \lambda\xi)(\mathbf{J}_\varphi\ \lambda\xi)^\top = \lambda^2 \mathbf{I}_{3\times 3}. \tag{2}$$

where $\xi \in C^2(\mathcal{T};\mathbb{R}^3)$ models the surface normal field. Note that ξ depends on φ, as it is a unitary vector orthogonal to the two columns of $\mathbf{J}_\varphi$. To summarize, SfT consists of finding the embedding φ and normal field ξ given the warp η, the flattening scale λ and the projection Π, by solving a nonlinear PDE system:

$$\text{Find } \varphi \in C^2(\mathcal{T};\mathbb{R}^3) \text{ st } \begin{cases} (\mathbf{J}_\varphi\ \lambda\xi)(\mathbf{J}_\varphi\ \lambda\xi)^\top = \lambda^2 \mathbf{I}_{3\times 3} & \text{Deformation} \\ \eta = \Pi \circ \varphi & \text{Reprojection.} \end{cases} \tag{3}$$

Eq. (3) involves first-order derivatives of the unknown function φ. Following [1], differentiating the reprojection constraint and substituting it into the deformation constraint yields:

$$\mathbf{J}_\eta \mathbf{J}_\eta^\top + \lambda^2 (\mathbf{J}_\Pi \circ \varphi)\xi\xi^\top (\mathbf{J}_\Pi \circ \varphi)^\top = \lambda^2 (\mathbf{J}_\Pi \circ \varphi)(\mathbf{J}_\Pi \circ \varphi)^\top, \tag{4}$$

where $\mathbf{J}_\Pi \circ \varphi$ is a 2×3 matrix that only depends on the surface depth φ_z. Eq. (4) is a PDE system of 3 independent equations in φ_z and ξ. Very recently, [2] obtained the pointwise solutions of Eq. (4), by ignoring the differential relationship between φ_z and ξ. Those solutions are called *non-holonomic* and [2] showed that they can be obtained analytically.

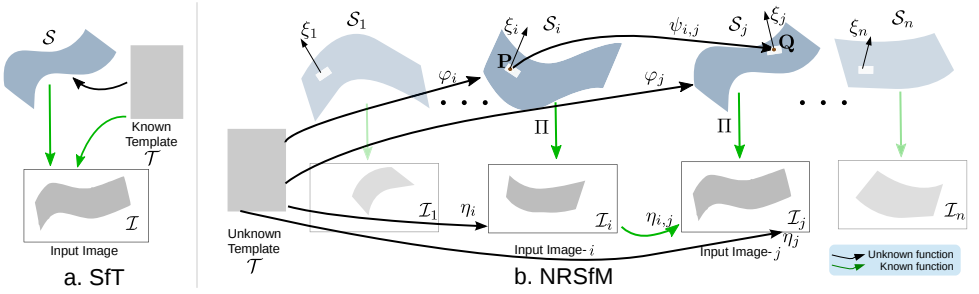

Figure 1: Geometric modeling of SfT and NRSfM.

3.2 From SfT to NRSfM: No Known Template, but More Images

To extend the first-order differential modeling of SfT to NRSfM we introduce n images showing a different deformation and keep the template $\mathcal{T}$ as an unknown. We use the index $i = 1,\dots,n$ to define the i-th shape $\mathcal{S}_i$, image $\mathcal{I}_i$, warp η_i and embedding φ_i. The inter-image registration warp $\eta_{i,j}$ is known and related to the unknown warps η_i and η_j as $\eta_{i,j} = \eta_j \circ \eta_i^{-1}$. We denote as $\psi_{i,j} \in C^2(\mathcal{S}_i;\mathcal{S}_j)$ the unknown isometric deformation between $\mathcal{S}_i$ and $\mathcal{S}_j$, where $\varphi_j = \psi_{i,j} \circ \varphi_i$. From Eq. (1), it is clear that $\psi_{i,j}$ preserves the first fundamental form between φ_i and φ_j:

$$\mathbf{J}_{\varphi_i}^\top \mathbf{J}_{\varphi_i} = \mathbf{J}_{\varphi_j}^\top \mathbf{J}_{\varphi_j} \tag{5}$$

In NRSfM the objective is to find the embeddings φ_i, $i = 1,\dots,n$ and the unknown template $\mathcal{T}$ given the pairwise image warps:

$$\text{Find} \left| \begin{array}{l} \mathcal{T} \subset \mathbb{R}^2 \\ \varphi_i \in C^2(\mathcal{T};\mathbb{R}^3) \\ i = 1,\dots,n \end{array} \right. \quad \text{st} \quad \left\{ \begin{array}{lll} \eta_{i,j} = \eta_j \circ \eta_i^{-1} \quad j = 1,\dots,n \quad j \neq i & & \text{Consistency} \\ \eta_i = \Pi \circ \varphi_i & & \text{Reprojection} \\ (\mathbf{J}_{\varphi_i} \lambda\xi_i)(\mathbf{J}_{\varphi_i} \lambda\xi_i)^\top = \lambda^2 \mathbf{I}_{3\times 3} & & \text{Deformation.} \end{array} \right. \tag{6}$$

3.3 Isometric NRSfM is not Locally Solvable at First-Order

We show that system (6) can be expressed as a nonlinear PDE system in terms of the surfaces' depth and normal, and the unknown template. Our main result is that the resulting system is not locally solvable, which means that its non-holonomic solutions are underconstrained. We first derive the NRSfM system for two views i and j. We start from Eq. (4), which combines the reprojection and deformation constraints of Eq. (6). We parametrize $\mathbf{J}_{\eta_i} \in C^2(\mathcal{T};\mathbb{R}^{2\times 2})$ in the following general form:

$$\mathbf{J}_{\eta_i} = \sigma \mathbf{M} \mathbf{R}_\theta \quad \text{with} \quad \mathbf{R}_\theta \mathbf{R}_\theta^\top = \mathbf{I}_{2\times 2} \quad \text{and} \quad \mathbf{M} = \begin{pmatrix} 1 & \beta \\ 0 & \alpha \end{pmatrix}, \tag{7}$$

where $\mathbf{R}_\theta$ is a 2D rotation of angle θ. Invoking Cholesky decomposition, the 4 dimensions of $\mathbb{R}^{2\times 2}$ are equivalent to $(\sigma,\theta,\beta,\alpha)$. $\mathbf{J}_{\eta_i}\mathbf{J}_{\eta_i}^\top = \sigma^2 \mathbf{M}\mathbf{M}^\top$ reveals that Eq. (4) is invariant to the 2D rotation $\mathbf{R}_\theta$. To use Eq. (4) with φ_j while following the consistency relation in Eq. (6), we differentiate $\eta_{i,j} = \eta_j \circ \eta_i^{-1}$ to $\left(\mathbf{J}_{\eta_j} \circ \eta_i^{-1}\right) \mathbf{J}_{\eta_i}^{-1} = \mathbf{J}_{\eta_{i,j}}$ and use Eq. (7) to obtain: $\mathbf{J}_{\eta_j} \circ \eta_i^{-1} = \sigma \mathbf{J}_{\eta_{i,j}} \mathbf{M} \mathbf{R}_\theta$. By multiplying each side to the right by its transpose, the rotation

vanishes:

$$\left(\mathbf{J}_{\eta_j}\circ\eta_i^{-1}\right)\left(\mathbf{J}_{\eta_j}\circ\eta_i^{-1}\right)^\top=\sigma^2\mathbf{J}_{\eta_{i,j}}\mathbf{M}\mathbf{M}^\top\mathbf{J}_{\eta_{i,j}}^\top. \tag{8}$$

As $\mathcal{T}$ is just required to be a conformal flattening of the 3D template we may choose the scale factor $\sigma=\lambda$. Introducing Eq. (7) and (8) in Eq. (4) we obtain the isometric NRSfM system of PDEs for two unknown surfaces φ_i and φ_j:

$$\begin{cases}\mathbf{M}\mathbf{M}^\top+(\mathbf{J}_\Pi\circ\varphi_1)\xi_i\xi_i^\top(\mathbf{J}_\Pi\circ\varphi_i)^\top=(\mathbf{J}_\Pi\circ\varphi_i)(\mathbf{J}_\Pi\circ\varphi_i)^\top\\ \mathbf{J}_{\eta_{i,j}}\mathbf{M}\mathbf{M}^\top\mathbf{J}_{\eta_{i,j}}^\top+(\mathbf{J}_\Pi\circ\varphi_j)\xi_j\xi_j^\top(\mathbf{J}_\Pi\circ\varphi_j)^\top=(\mathbf{J}_\Pi\circ\varphi_j)(\mathbf{J}_\Pi\circ\varphi_j)^\top.\end{cases} \tag{9}$$

Eq. (9) is an algebraic system of 6 equations and 8 unknowns $(\varphi_{i,z},\xi_i,\varphi_{j,z},\xi_j,\alpha,\beta)$ at every point. The non-holonomic solutions of system (9) are thus underconstrained for two views. In the general case of n views the system has $3n+2$ unknowns and $3n$ independent equations. Its non-holonomic solutions are thus underconstrained for $n>2$ views as well. Our main result is that without further assumptions, one cannot solve isometric NRSfM by relaxing the relationship between depth and normal, as was done in SfT [1, 2].

4 Infinitesimally Planar Isometric NRSfM

We show isometric NRSfM can be solved locally (and analytically) if we assume that the surfaces are infinitesimally planar. This approximation is equivalent to representing the surfaces with triangular meshes, where the size of the triangles is infinitesimally small. The result is that higher order surface derivatives are *locally* zero.

4.1 Infinitesimal Projective Structure

We define $\hat{\varphi}_i$ as the locally planar approximation of the embedding φ_i:

$$\hat{\varphi}_i=\varphi_i+\mathbf{J}_{\varphi_i}\delta, \tag{10}$$

where $\delta\in\mathbb{R}^2$ are local 2D coordinates around each point in $\mathcal{T}$. Eq. (10) parametrizes the tangent planes of $\mathcal{S}_i$. We show next that two corresponding tangent planes on $\mathcal{S}_i$ and $\mathcal{S}_j$ are related by a rigid transform when $\psi_{i,j}$ is an isometry.

Differentiating $\varphi_j=\psi_{i,j}\circ\varphi_i$ gives $\mathbf{J}_{\varphi_j}=(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)\mathbf{J}_{\varphi_i}$. Using Eq. (5) we show that the 3×3 matrix $\left(\mathbf{J}_\psi\circ\varphi_i\right)$ is indeed orthonormal:

$$\mathbf{J}_{\varphi_j}^\top\mathbf{J}_{\varphi_j}=\mathbf{J}_{\varphi_i}^\top(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)^\top(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)\mathbf{J}_{\varphi_i}=\mathbf{J}_{\varphi_i}^\top\mathbf{J}_{\varphi_i}\quad\Longrightarrow\quad(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)^\top(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)=\mathbf{I}_{3\times3}. \tag{11}$$

Using Eq. (11) we represent $\hat{\varphi}_j$ as a rigid transformation of $\hat{\varphi}_i$:

$$\hat{\varphi}_j=\varphi_j+\mathbf{J}_{\varphi_j}\delta=\psi_{i,j}\circ\varphi_i+(\mathbf{J}_{\psi_{i,j}}\circ\varphi_i)\mathbf{J}_{\varphi_i}\delta=\mathbf{t}_{ij}+\mathbf{R}_{ij}\hat{\varphi}_i. \tag{12}$$

where $\mathbf{R}_{ij}=\mathbf{J}_{\psi_{i,j}}\circ\varphi_i$ is a 3D rotation from Eq. (11) and $\mathbf{t}_{ij}=\psi_{i,j}\circ\varphi_i-\mathbf{R}_{ij}\varphi_i=\varphi_j-\mathbf{R}_{ij}\varphi_i$ represents a translation. Eq. (12) means that two corresponding tangent planes in $\mathcal{S}_i$ and $\mathcal{S}_j$ are related by a rigid transform.

We modify the reprojection constraint in Eq. (6) for an infinitesimally planar surface as $\hat{\eta}_i=\Pi\circ\hat{\varphi}_i$. $\hat{\eta}_i$ as a function of δ is the warp between the template and the projection of the tangent plane. Using homogeneous coordinates we have $\begin{pmatrix}\hat{\eta}_i\\1\end{pmatrix}\propto\hat{\varphi}_i$ and from Eq. (10)

$\begin{pmatrix}\hat{\eta}_i\\1\end{pmatrix} \propto \mathbf{H}_i \begin{pmatrix}\delta\\1\end{pmatrix}$, with $\mathbf{H}_i = (\mathbf{J}_{\varphi_i}\ \varphi_i)$. The warp $\hat{\eta}_i$ is thus a homography induced by the tangent plane. The image warp $\hat{\eta}_{i,j} = \hat{\eta}_j \circ \hat{\eta}_i^{-1}$ too is thus a homography given by $\mathbf{H}_{i,j} = \mathbf{H}_j\mathbf{H}_i^{-1}$.

The structure of $\mathbf{H}_{i,j}$ is well-known [11]: it represents the transformation between two images showing the projection of two planes related by a rigid transform (the above-derived $\mathbf{R}_{ij}$ and $\mathbf{t}_{ij}$). From [11], $\mathbf{H}_{i,j}$ can be decomposed as:

$$\mathbf{H}_{i,j} = \mathbf{R}_{ij} + \xi_i \mathbf{t}_{ij}^\top, \quad \text{where} \quad \mathbf{R}_{ij} = \mathbf{J}_\psi \circ \varphi_i \quad \text{and} \quad \mathbf{t}_{ij} = \varphi_j - \mathbf{R}_{ij}\varphi_i. \tag{13}$$

Given $\mathbf{H}_{i,j}$, we can thus extract the normal field of the surface and φ_i, However, [11] shows that there is always a two-fold solution for ξ_i, $\mathbf{R}_{ij}$ and $\mathbf{t}_{ij}$. With two views it is thus not possible to disambiguate reconstruction. Extra cues must be introduced. [18] proposes to use smoothness but it is not guaranteed to give the correct solution. If we use three or more views we get a collection of normals for each point (*i.e.* two for each pair of views). We can thus disambiguate the normals using more than 2 views and clustering the normals to find an agreement with the dot-product measure. A more detailed explanation is given in §4.3.

4.2 Differential Homography Computation

We now show how to obtain $\mathbf{H}_{i,j}$ from the registration warp $\eta_{i,j}$. Given a point $\mathbf{p} \in \mathcal{I}_i$, we assume that $\eta_{i,j}(\mathbf{p}+\varepsilon) = \hat{\eta}_{i,j}(\varepsilon)$ for a small $\varepsilon = (\varepsilon_u\ \varepsilon_v)^\top$, which gives:

$$\begin{pmatrix}\rho(\varepsilon)\eta_{i,j}(\mathbf{p}+\varepsilon)\\ \rho(\varepsilon)\end{pmatrix} = \mathbf{H}_{i,j}\begin{pmatrix}\varepsilon\\1\end{pmatrix} \quad \text{where} \quad \mathbf{H}_{i,j} = \begin{pmatrix}a & b & c\\ g & h & k\\ d & e & 1\end{pmatrix}, \tag{14}$$

and $\rho(\varepsilon) = d\varepsilon_x + e\varepsilon_y + 1$ is an unknown linear function. When $\varepsilon = \mathbf{0}$ then $\rho = 1$ and $\eta_{i,j}(\mathbf{p}) = (c\ \ k)^\top$, from which we obtain c and k. By taking first and second derivatives with respect to ε on both sides of Eq. (14) and evaluating them at $\varepsilon = \mathbf{0}$ we obtain the following system of equations in the elements of $\mathbf{H}_{i,j}$:

$$\eta_{i,j} = \begin{pmatrix}c\\k\end{pmatrix} \quad \mathbf{J}_{\eta_{i,j}} = \begin{pmatrix}a-cd & b-ce\\ g-kd & h-ke\end{pmatrix} \quad \frac{\partial^2\eta_{i,j}}{\partial u^2} = \begin{pmatrix}-2d(a-cd)\\-2d(g-kd)\end{pmatrix}$$
$$\frac{\partial^2\eta_{i,j}}{\partial v^2} = \begin{pmatrix}-2e(b-ce)\\-2e(h-ke)\end{pmatrix} \quad \frac{\partial^2\eta_{i,j}}{\partial u\partial v} = \begin{pmatrix}-2d(b-ce)-ae\\-2d(h-ke)-ge\end{pmatrix}. \tag{15}$$

System (15) gives 12 equations for 8 unknowns as all derivatives of η are known. We solve it using linear least-squares.

4.3 Algorithm

Our algorithm involves the following steps given n views of the surface: *1)* Select one view as the reference and compute the registration warp with respect to all other views using e.g. [20]. *2)* For every point in the reference view and every possible pair ($n-1$ pairs) we obtain a homography using first and second order derivatives of the registration warp (Eq. (15)). *3)* Decompose the $n-1$ homographies that we obtain for each point between the reference image to the others. Thus we have two normals from each homography at this step.

4) Remove normals that are not front facing. *5)* Cluster the normals and obtain two normals corresponding to the two largest clusters: if the two largest clusters are similarly supported, then disambiguate using agreement with neighbours (*e.g.* smoothness). Otherwise keep the normal of the largest cluster. *6)* Integrate the normal field to obtain the reference surface embedding. *7)* To reconstruct the $(n-1)$ remaining surfaces we can either change the reference surface or use SfT given that the surface computed for the reference image is now the known template.

Even though 3 views is the minimal case, in practice we use more views to avoid ambiguities due to the presence of noise and deformations. Note that we can use any image as the reference image for each point.

5 Experimental Evaluation

We tested our method with synthetic data along with two real datasets of a deforming piece of paper and cloth. The different views show large deformations and wide baseline viewpoints. We computed image warps using SIFT keypoints [10] followed by robust registration [14]. We modeled the inter-image warps with Bicubic B-Splines (BBS) with 20×20 control points [4]. We compared our method (*DiffH*) with four other: *DiscH* is our pipeline with discrete homography computation from 4 point correspondences hallucinated using η at a distance r from the central point, *Varol09* [18], *Taylor10* [16] and *Vicente12* [19]. The comparison was done for various numbers of views and noise levels for the synthetic data and for different numbers of views for the real datasets. Quantitative evaluations were obtained by measuring the shape error (mean error of the computed normals in degrees) and the depth error (mean error in the reconstructed 3D coordinates).

5.1 Synthetic Data

We simulated 10 different scenes of an isometrically deformed sheet of paper [13]. The images were taken at a focal length of 200 *px* and their dimensions are 640 $px \times 480$ *px*. We randomly selected 400 correspondences computed with a Gaussian noise of standard deviation σ in *px*. We varied the number of views from $n = 4$ to $n = 10$ and the noise standard deviation from $\sigma = 0$ to $\sigma = 4$ *px*. We fixed $n = 10$ for the evaluation in varying noise and $\sigma = 1.2$ *px* for the evaluation in varying number of views. The results are shown in the top row of figure 2.

The results show that *Taylor10* and *Vicente12* do not produce correct reconstructions with the shape error around 80 degrees. The reason for this is for the most part, the perspective nature of the images. Both *Taylor10* and *Vicente12* methods use the orthographic camera. The depth errors for these two methods are not shown as they go beyond the scale used in the graphs. Using 10 views and no added noise, we observed a depth error of 194.14 *mm* for *Taylor10* and 209.2 *mm* for *Vicente12*. *Varol09* also failed to produce good results as its approach to normal disambiguation using solely smoothness is too weak. The shape error for *DiffH* on the other hand, remains lower than 10 degrees for $n > 4$. *DiscH* follows behind with shape error about 4 degrees larger than that for *DiffH*. Clearly *DiscH* is able to reconstruct the surfaces in most circumstances but we observe better reconstructions with *DiffH* owing to the more stable local homographies. As *DiscH* estimates the homographies using a radius parameter that determines how many points or the area of the object are considered for a single homography, the accuracy of the result for a particular value of the radius parameter

depends heavily on the deformation. We use the optimal value for r in *DiscH*, while we also observed that at the limit with very small values of r the reconstructions were worse.

5.2 The Real Datasets

We have constructed two different real datasets. The first shows a sheet of paper (the Hulk dataset) and the second shows a T-shirt (the T-shirt dataset). The Hulk dataset consists of a set of 10 images taken at different unrelated smooth deformations. We use the cover of a comics as texture. The image size is 4928 $px \times 3264$ px with a focal length of 3800 px. The T-shirt dataset also consists of a set of 10 images taken at different deformations. The image size is 4800 $px \times 3200$ px with a focal length of 3800 px. We used SfM using several images to compute the ground truth 3D shape for both of these datasets.

We evaluated the five different methods with varying number of views. The results are shown in the bottom row of figure 2 and confirm our observations for the synthetic data. Again the depth errors are shown only for the three methods: *DiffH*, *DiscH* and *Varol09*. With 10 views, the mean depth error for *Taylor10* is 48.4 *mm* and for *Vicente12* it is 86.5 *mm* in the Hulk dataset, and in the T-shirt dataset they are 47.3 *mm* and 76.9 *mm* respectively.

We also show the texture mapped reconstructions obtained using 10 views for all the compared methods on three simple examples for each real dataset in figure 3. The results show that *Taylor10* and *Vicente12* miscalculated the flips and thus reconstructed the wrong shape because they use the orthographic camera. *Varol09* does not disambiguate the normals properly in most cases, thus producing good shape only for some parts of the object or for some specific deformations. These observations can also be confirmed by the shape and depth error measurements given below each reconstruction.

6 Conclusions

We have presented in this paper the first differential modeling and study of isometric NRSfM. Our model unifies SfT and NRSfM and shows that non-holonomic solutions in NRSfM are underconstrained: the relationship between depth and normal cannot be directly relaxed as in SfT. We have shown however that the problem has a solution for $n > 2$ views if we consider the surface embeddings to be infinitesimally planar. Our solution involves only linear least-squares and analytical homography decompositions. We have given a method to deal with ambiguities in the general case of n views. We showed that our method clearly outperforms the state of the art and produces very accurate results in sparse unordered datasets that show wide-baseline viewpoints and large deformations.

Acknowledgements. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE.

References

[1] A. Bartoli and T. Collins. Template-based isometric deformable 3D reconstruction with sampling-based focal length self-calibration. In *CVPR*, 2013.

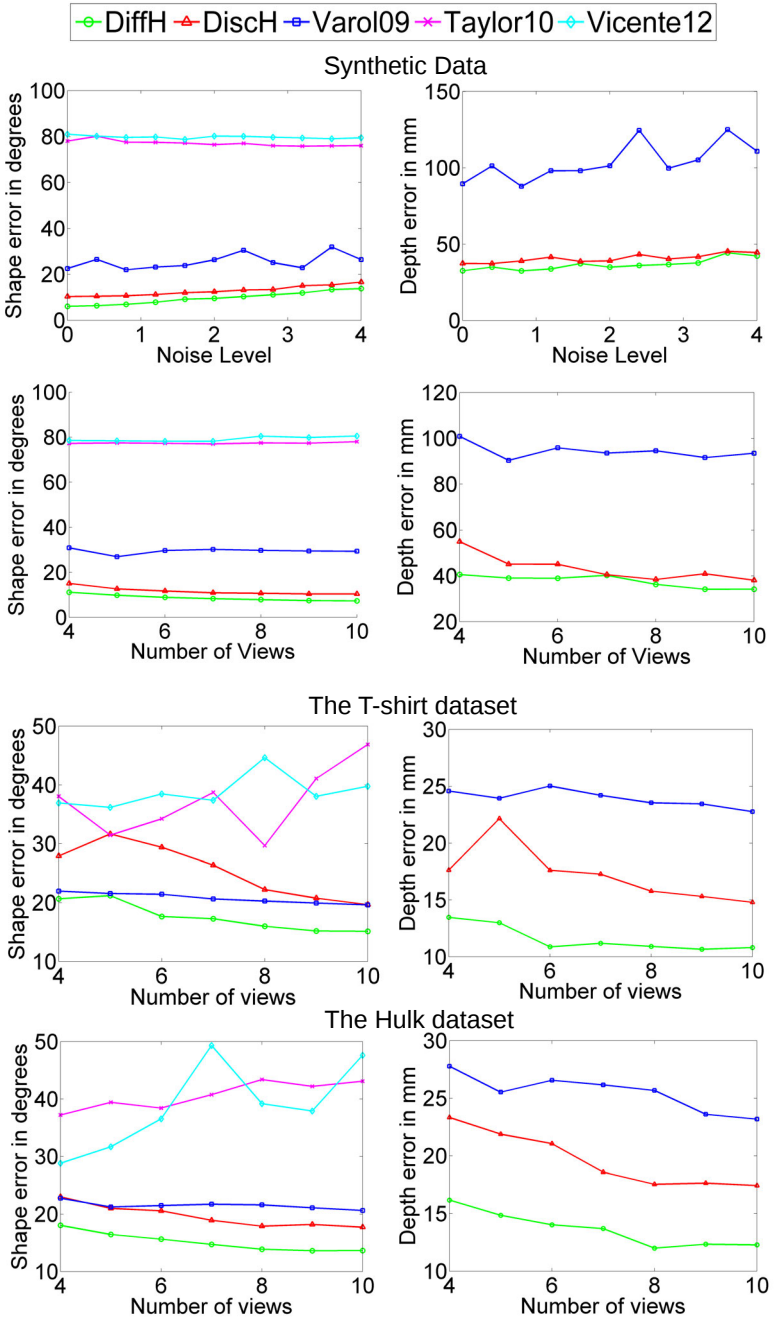

Figure 2: Error plots for synthetic data and the real datasets.

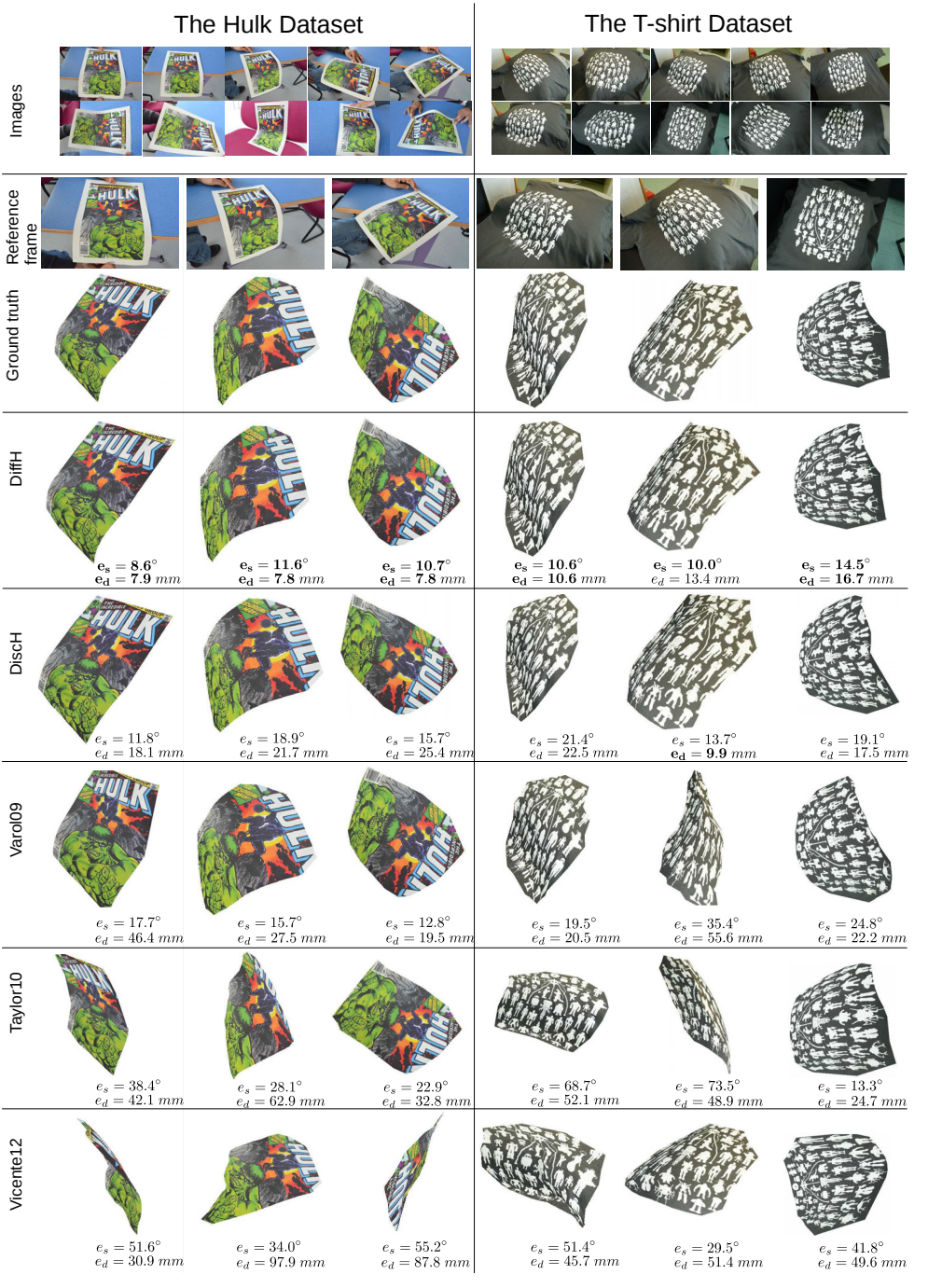

Figure 3: Qualitative results for three examples in the Hulk dataset and the T-shirt dataset: e_s is the shape error in degrees and e_d is the depth error in *mm*.

[2] A. Bartoli, Y. Gerard, F. Chadebecq, and T. Collins. On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces. In *CVPR*, 2012.

[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000.

[4] F. Brunet, V. Gay-Bellile, A. Bartoli, N. Navab, and R. Malgouyres. Feature-driven direct non-rigid image registration. *International Journal of Computer Vision*, 93(1): 33–52, 2011.

[5] T. Collins and A. Bartoli. Locally affine and planar deformable surface reconstruction from video. In *International Workshop on Vision, Modeling and Visualization*, 2010.

[6] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *CVPR*, 2012.

[7] A. Del Bue, F. Smeraldi, and L. Agapito. Non-rigid structure from motion using non-parametric tracking and non-linear optimization. In *CVPRW*, 2004.

[8] P. F.U. Gotardo and A. M. Martinez. Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(10):2051–2065, 2011.

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] E. Malis and M. Vargas. Deeper understanding of the homography decomposition for vision-based control. Rapport de recherche RR-6303, INRIA, 2007. URL http://hal.inria.fr/inria-00174036.

[12] A. Malti, R. Hartley, A. Bartoli, and J.-H. Kim. Monocular template-based 3D reconstruction of extensible surfaces with local linear elasticity. In *CVPR*, 2013.

[13] M. Perriollat and A. Bartoli. A computational model of bounded developable surfaces with application to image-based three-dimensional reconstruction. *Journal of Visualization and Computer Animation*, 24(5):459–476, 2013.

[14] D. Pizarro and A. Bartoli. Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 97(1):54–70, 2012.

[15] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(5): 931–944, 2011.

[16] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010.

[17] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008.

[18] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *CVPR*, 2009.

[19] S. Vicente and L. Agapito. Soft inextensibility constraints for template-free non-rigid reconstruction. In *ECCV*, 2012.

[20] G. Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.