# Isometric Non-Rigid Shape-from-Motion in Linear Time

Shaifali Parashar[1], Daniel Pizarro[2,1] and Adrien Bartoli[1]
[1]ALCoV-ISIT, UMR 6284 CNRS / Université d'Auvergne, Clermont-Ferrand, France
[2]Geintra Research Group, Universidad de Alcalá, Alcalá de Henares, Spain
shaifali.parashar@gmail.com,dani.pizarro@gmail.com,adrien.bartoli@gmail.com

## Abstract

*We study Isometric Non-Rigid Shape-from-Motion (Iso-NRSfM): given multiple intrinsically calibrated monocular images, we want to reconstruct the time-varying 3D shape of an object undergoing isometric deformations. We show that Iso-NRSfM is solvable from the warps (the inter-image geometric transformations). We propose a new theoretical framework based on Riemmanian manifolds to represent the unknown 3D surfaces, as embeddings of the camera's retinal planes. This allows us to use the manifolds' metric tensor and Christoffel Symbol fields, which we prove are related across images by simple rules depending only on the warps. This forms a set of important theoretical results. Using the infinitesimal planarity formulation, it then allows us to derive a system of two quartics in two variables for each image pair. The sum-of-squares of these polynomials is independent of the number of images and can be solved globally, forming a well-posed problem for $N \geq 3$ images, whose solution directly leads to the surface's normal field. The proposed method outperforms existing work in terms of accuracy and computation cost on synthetic and real datasets.*

## 1. Introduction

One of the main problems in 3D vision is to reconstruct an object's 3D shape from multiple views. This has been solved for the specific case of rigid objects from inter-image visual motion, and is known as Shape-from-Motion (SfM) [13]. However, SfM breaks down for non-rigid objects. Two ways to exploit visual motion for non-rigid object reconstruction have been proposed: Shape-from-Template (SfT) [2, 3, 26, 5] and Non-Rigid Shape-from-Motion (NRSfM) [4, 17, 1, 16, 25]. The latter is a direct extension of SfM to the non-rigid case. The former however, is not. Indeed, the inputs of SfT are a single image and the object's template, and its output is the object's deformed shape. The template is a very strong object-specific prior, as it includes a reference shape, a texture map and a deformation model. Most SfT methods use physics-based deformation models such as isometry [3, 26]. This is because isometry is a very good approximation to the deformation of many real objects. The inputs of NRSfM are multiple images and its output is the object's 3D shape for every image. In NRSfM, the rigidity constraint of SfM is replaced by constraints on the object's deformation model. NRSfM methods were proposed for a variety of deformation models: the low-rank shape basis [17], the trajectory basis [20, 8], isometry [4, 25] and elasticity [7]. Existing methods suffer one or several limitations amongst solution ambiguities, low accuracy, ill-posedness, inability to handle missing data and high computation cost. NRSfM thus still exists as an open research problem.

We present a solution to NRSfM with the isometric deformation model, that we hereinafter denote Iso-NRSfM. We model Iso-NRSfM using concepts from Riemannian geometry. Our framework relates the 3D shape to the inter-image warps, which we simply call warps. These may be computed from keypoint correspondences in several ways [10, 22], and we assume they are known. More specifically, we model the object's 3D shape for each image by a Riemmanian manifold and deformations as isometric mappings. We parameterize each manifold by embedding the corresponding retinal plane. This allows us to reason on advanced surface properties, namely the metric tensor and Christoffel Symbols, directly in retinal coordinates, and in relationship to the warps. We prove two new theorems showing that the metric tensor and Christoffel Symbols may be transferred between views using only the warp. For an infinitesimally planar surface, we obtain a system of two quartics in two variables that involves up to second order derivatives of the warps. This system holds at each point. Its solution gives an estimate of the metric tensor, and thus of the surface's normal, in all views. The shape is finally recovered by integrating the normal fields for each view.

The proposed method has the following features. *1)* It has a linear complexity in the number of views and number of points. *2)* It uses a well-posed point-wise solution from $N \geq 3$ views, thus covering the minimum data case.

*3)* It naturally handles missing data created by occlusions. *4)* It substantially outperforms existing methods in terms of complexity and accuracy, as we experimentally verified using synthetic and real datasets. Beyond the proposed method, we bring a completely new theoretical framework to NRSfM allowing one to exploit the surface's metric tensor and Christoffel Symbols in a simple and neat way.

## 2. State-of-the-Art

Existing NRSfM methods can be divided into three main categories: *i)* object-wise, *ii)* piece-wise and *iii)* point-wise methods. *i)* solves for the entire object's shape at once. This group includes methods that assume a low-dimensional space of deformed shapes [17, 11, 12]. These methods have been extensively studied in the recent years with the low-rank prior [11] and other constraints such as temporal smoothness [17] or point trajectory constraints [20, 8]. They have demonstrated to be accurate for objects with a low number of deformation modes, such as a talking face and articulated objects. These methods suffer in the presence of missing data and may present ambiguities [28] (for instance due to the orthographic camera assumption). *i)* also includes methods using physics-based models such as isometry [25], elastic deformations [7] or particle-based interactions [6]. [25] copes with missing data but involves costly non-convex optimization, which requires a very good initialization. Recently [7, 6] proposed a sequential solution based on elasticity [7] and particle interaction [6]. Both methods are promising but require rigid motion at the beginning of the sequence to reconstruct the object's shape using rigid SfM. Those methods are related to SfT.

Methods in *ii)* and *iii)* are sometimes also called local methods. In piece-wise methods one selects a simple model that approximates the shape of a small region of the surface. NRSfM is then solved for each region. This can be analytical for planes [1] and local rigid motions with both the orthographic [16] and perspective [24] cameras. More complex models, such as the local quadratic models [15], require non-linear iterative optimization. After solving for each region's approximate shape, a second step is to stitch together all reconstructions, imposing some order of continuity in the surface. In [23] stitching is done using submodular optimization. For some other models stitching can be solved by Linear Least Squares (LLS) [4]. Piece-wise methods are very problematic due to the need for segmenting the image domain in regions from which the local models are computed. Region segmentation is costly and difficult to define optimally for general surfaces. This has a major impact in the efficiency and accuracy of these methods.

Point-wise methods replace local regions with infinitesimal regions, which allows one to describe NRSfM as a system of Partial Differential Equations (PDE) involving differential properties of the shape and derivatives of the warps [10]. [4] presents a point-wise solution for isometric NRSfM assuming that the surface is infinitesimally planar. It gives analytical solutions to compute the surface's twofold ambiguous normal at any point from a pair of views. The strategy given in [4] is to average over the normals that are compatible across different pairs. This finds a single normal per point but requires in practice a large amount of image pairs to be accurate and may thus be very expensive. The global shape is then obtained by integration of the normal fields by means of LLS. Although [4] reports better results with respect to other methods, we show that it fails in several other cases.

Point-wise methods form a promising solution for Iso-NRSfM. In principle, they allow one to overcome the complexity, missing data, and accuracy limitations of other methods. However, in practice, no theoretical framework and practical method were proposed which overcome these limitations. Our paper attempts to fill this gap by proposing a Riemannian framework coupled with infinitesimal planarity, leading to a method solving NRSfM accurately, using small globally solvable optimization problems, and in time complexity linear in the number of images and points.

## 3. Mathematical Model

### 3.1. General Model

Our model of NRSfM is shown in figure 1. We have $N$ input images $\mathcal{I}_1, \dots, \mathcal{I}_N$ that show the projection of different isometric deformations of the same surface. The registration between the pair of images $\mathcal{I}_i$ and $\mathcal{I}_j$ is known and denoted by the functions $\eta_{ij}$ and $\eta_{ji}$, called warps. In practice, we compute them from keypoint correspondences using [22]. Abusing notation, we also use $\mathcal{I}_i$ to denote an image's retinal plane, with $\mathcal{I}_i \subset \mathbb{R}^2$. Surfaces in 3D are

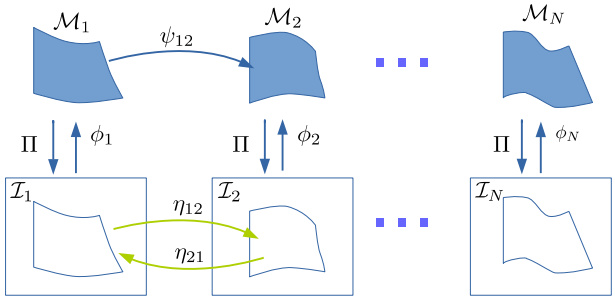

Figure 1: The proposed model of NRSfM, where each surface is a Riemannian manifold.

assumed to be Riemannian manifolds. This allows us to define lengths, angles and tangent planes on the surface [18]. We denote $\mathcal{M}_i$ as the $i$th manifold, which can be seen as a two-dimensional subset embedded in 3D, $\mathcal{M}_i \subset \mathbb{R}^3$. We use the extrinsic definition of $\mathcal{M}_i$, where a function embeds a subset of the plane $\mathbb{R}^2$ into $\mathbb{R}^3$. With embedding functions,

one can easily compute manifold characteristics [19] such as metric tensors and Christoffel Symbols. However, these characteristics change according to the coordinate frame. We use the retinal plane $\mathcal{I}_i$ as coordinate frame for $\mathcal{M}_i$ and define $\phi_i \in C^\infty(\mathcal{I}_i, \mathbb{R}^3)$ as the *image embedding* for $\mathcal{M}_i$. We define $\psi_{ij}$ as the isometric mapping between manifolds $\mathcal{M}_i$ and $\mathcal{M}_j$.

### 3.2. Image Embeddings

We use the perspective camera as projection model. For a 3D point $\mathbf{z} = \begin{pmatrix} z^1 & z^2 & z^3 \end{pmatrix}^\top$ we define perspective projection $\Pi$ as:

$$\mathbf{x} = \Pi(\mathbf{z}) \qquad \mathbf{x} = \begin{pmatrix} \frac{z^1}{z^3} & \frac{z^2}{z^3} \end{pmatrix}^\top, \tag{1}$$

where $\mathbf{x}$ is the projected point's retinal coordinates. The image embeddings $\phi_i$ with $i = 1, \dots, N$ define the inverse of a perspective projection, as they map retinal coordinates to a 3D surface. They thus satisfy the following identity:

$$\mathbf{x} = (\Pi \circ \phi_i)(\mathbf{x}) \qquad i = 1, \dots, N. \tag{2}$$

Smooth functions that comply with (2) can be expressed with a depth map $\rho_i \in C^\infty(\mathcal{I}_i, \mathbb{R})$, where:

$$\phi_i(\mathbf{x}) = \rho_i(\mathbf{x}) \begin{pmatrix} \mathbf{x} & 1 \end{pmatrix}^\top \qquad i = 1, \dots, N. \tag{3}$$

### 3.3. Metric Tensors

The metric tensor [18] of $\phi_i$ is denoted $\mathbf{g}_{mn}[\phi_i]$. We use the standard Einstein's tensor notation and thus $\mathbf{g}_{mn}[\phi_i]$ is a combined reference to all elements of the metric tensor, a $2 \times 2$ matrix in this case. The indexes $m$ and $n$ reference to each component of the coordinate frame of $\phi_i$, that we denote with $\mathbf{x} = \begin{pmatrix} x^1 & x^2 \end{pmatrix}^\top$. We have $\mathbf{z} = \phi_i(\mathbf{x})$, where $\mathbf{z} = \begin{pmatrix} z^1 & z^2 & z^3 \end{pmatrix}^\top$. The metric tensor of $\phi_i$ is then:

$$\mathbf{g}_{mn}[\phi_i] = \frac{\partial z^s}{\partial x^m} \frac{\partial z^k}{\partial x^n} \delta_{sk}, \tag{4}$$

with $\delta_{sk}$ Kronecker's delta function. We remind that the summation in (4) is done over indices $s$ and $k$. The inverse of the metric tensor is expressed with raised indexes $\mathbf{g}^{mn}[\phi_i]$. Given the change of coordinates:

$$\mathbf{x} = \eta(\mathbf{y}), \quad \text{with} \quad \mathbf{y} = \begin{pmatrix} y^1 & y^2 \end{pmatrix}^\top, \tag{5}$$

the metric tensor of $\phi_i \circ \eta$ is obtained as:

$$\mathbf{g}_{st}[\phi_i \circ \eta] = \frac{\partial x^m}{\partial y^s} \frac{\partial x^n}{\partial y^t} \mathbf{g}_{mn}[\phi_i], \tag{6}$$

where in (6) we omit that $\mathbf{g}_{mn}[\phi_i]$ is composed with $\eta$ to simplify notation.

We introduce next a theorem regarding the relationship between the metric tensors in the different manifolds $\phi_i$ with $i = 1, \dots, N$ if the mappings $\psi_{ij}$ with $\{i, j\} \in \{1, \dots, N\}^2$ are isometric. This theorem is fundamental for the formulation of our method.

**Theorem 1.** *Let $\psi_{ij}$ be an isometric mapping between the manifolds $\mathcal{M}_i$ and $\mathcal{M}_j$ describing Iso-NRSfM, then $\mathbf{g}_{mn}[\phi_j] = \mathbf{g}_{mn}[\phi_i \circ \eta_{ji}]$ with $(i, j) \in \{1, \dots, N\}^2$.*

*Proof.* We first write $\phi_j$ in terms of $\phi_i$ using the isometric mapping $\psi_{ij}$:

$$\phi_j = \psi_{ij} \circ \phi_i \circ \eta_{ji}. \tag{7}$$

From (6) and (7) we have:

$$\mathbf{g}_{mn}[\phi_j] = \mathbf{g}_{mn}[(\psi_{ij} \circ \phi_i) \circ \eta_{ji}] = \frac{\partial x^s}{\partial y^m} \frac{\partial x^t}{\partial y^n} \mathbf{g}_{st}[\psi_{ij} \circ \phi_i]. \tag{8}$$

By definition isometric mappings do not change the local metric and so $\mathbf{g}[\psi_{ij} \circ \phi_i] = \mathbf{g}[\phi_i]$, which applied to (8) gives:

$$\mathbf{g}_{mn}[\phi_j] = \frac{\partial x^s}{\partial y^m} \frac{\partial x^t}{\partial y^n} \mathbf{g}_{st}[\phi_i]. \tag{9}$$

Identifying (6) with (9) gives the sought equality $\mathbf{g}_{mn}[\phi_j] = \mathbf{g}_{mn}[\phi_i \circ \eta_{ji}]$.

□

Theorem 1 shows that given $\mathbf{g}_{mn}[\phi_i]$, we can find the metric tensor $\mathbf{g}_{mn}[\phi_j]$ in the manifold $\mathcal{M}_j$ by making a change of variable with the known function $\eta_{ji}$. Note that this is not true in general for non-isometric mappings.

### 3.4. Christoffel Symbols

The Christoffel Symbols (CS) [18] of the second kind are functional arrays that describe several properties of a Riemannian manifold, such as the curvature tensor, the geodesic equations of curves and the parallel transport of vectors in surfaces. We denote the CS of function $\phi_i$ as $\mathbf{\Gamma}^p_{mn}[\phi_i]$. Sometimes it is useful to represent the CS of $\phi_i$ as two $2 \times 2$ matrices $\mathbf{\Gamma}^1_{mn}[\phi_i]$ and $\mathbf{\Gamma}^2_{mn}[\phi_i]$, where 1 and 2 are bases of the coordinate frame of $\phi_i$. The CS are obtained from the metric tensor and its first derivatives:

$$\mathbf{\Gamma}^p_{mn}[\phi_i] = \frac{1}{2} \mathbf{g}^{pl}[\phi_i] \left( \mathbf{g}_{lm,n}[\phi_i] + \mathbf{g}_{ln,m}[\phi_i] - \mathbf{g}_{mn,l}[\phi_i] \right), \tag{10}$$

where $\mathbf{g}_{lm,n} = \partial_n \mathbf{g}_{lm}$. Given a change of coordinates $\mathbf{x} = \eta(\mathbf{y})$, the CS in the new coordinates are given as:

$$\mathbf{\Gamma}^q_{st}[\phi_i \circ \eta] = \frac{\partial x^m}{\partial y^s} \frac{\partial x^n}{\partial y^t} \mathbf{\Gamma}^p_{mn}[\phi_i] \frac{\partial y^q}{\partial x^p} + \frac{\partial y^q}{\partial x^l} \frac{\partial^2 x^l}{\partial y^s \partial y^t}. \tag{11}$$

Note that although the CS are described using tensorial notation, they are not tensors and thus (11) does not correspond to the way tensors change coordinates. We now give a corollary of Theorem 1 regarding the relationship of the CS between the manifolds in Iso-NRSfM.

**Corollary 1.** *Let $\psi_{ij}$ be the isometric mapping between the manifolds $\mathcal{M}_i$ and $\mathcal{M}_j$ describing Iso-NRSfM, then $\mathbf{\Gamma}^p_{mn}[\phi_j] = \mathbf{\Gamma}^p_{mn}[\phi_i \circ \eta_{ji}]$ with $(i,j) \in \{1,\dots,N\}^2$.*

*Proof.* As described in (10), $\mathbf{\Gamma}^p_{mn}[\phi_j]$ is a function of $\mathbf{g}_{mn}[\phi_j]$ and its derivatives. From Theorem 1 we have that $\mathbf{g}_{mn}[\phi_j] = \mathbf{g}_{mn}[\phi_i \circ \eta_{ji}]$. By multiplying this expression in both sides by $\mathbf{g}^{mn}[\phi_j]$ we have:

$$\mathbf{g}^{mn}[\phi_j]\mathbf{g}_{mn}[\phi_j] = \mathbf{g}^{mn}[\phi_j]\mathbf{g}_{mn}[\phi_i \circ \eta_{ji}] = \delta_{mn}, \quad (12)$$

from which we deduce that $\mathbf{g}^{mn}[\phi_j] = \mathbf{g}^{mn}[\phi_i \circ \eta_{ji}]$. Also, by differentiating both sides we obtain

$$\partial_l \mathbf{g}_{mn}[\phi_j] = \partial_l \mathbf{g}_{mn}[\phi_i \circ \eta_{ji}], \quad (13)$$

obtaining $\mathbf{g}_{mn,l}[\phi_j] = \mathbf{g}_{mn,l}[\phi_i \circ \eta]$. By substitution of these identities in (10) we obtain:

$$\mathbf{\Gamma}^p_{mn}[\phi_j] = \frac{1}{2}\mathbf{g}^{pl}[\phi_i \circ \eta_{ji}]\left(\mathbf{g}_{lm,n}[\phi_i \circ \eta_{ji}] + \mathbf{g}_{ln,m}[\phi_i \circ \eta_{ji}] - \mathbf{g}_{mn,l}[\phi_i \circ \eta_{ji}]\right), \quad (14)$$

and thus the equality $\mathbf{\Gamma}^p_{mn}[\phi_j] = \mathbf{\Gamma}^p_{mn}[\phi_i \circ \eta_{ji}]$ holds. □

This corollary has a similar interpretation as Theorem 1. If the CS are known in one manifold, isometric mappings allow one to reconstruct them in the other manifolds via a change of variable given by the warps.

### 3.5. Infinitesimal Planarity

In infinitesimal planarity one assumes that a surface is at each point approximately planar. This is fundamentally different from piece-wise planarity: in infinitesimal planarity, the surface is globally curved and represented infinitesimally by an infinite set of planes. In other words, each infinitesimal model agrees with the global surface at the point where infinitesimal planarity is used only at zeroth order. We use this approximation to find a point-wise solutions to Iso-NRSfM. We proceed by assuming that any $\mathcal{M}_i$ for $i \in \{1\dots,N\}$ is a plane and deriving the differential properties of the image embedding $\phi_i$, the metric tensor and the CS. We use these differential properties at each point of $\mathcal{M}_i$.

We give two theorems and a corollary about the special properties of $\mathcal{M}_i$ with $i \in \{1,\dots,N\}$, assuming planarity: *1)* Theorem 2 shows that the inverse depth in the embedding $\phi_i$ is a linear function, *2)* Corollary 2 states that point-wise both the metric tensor and the CS of $\mathcal{M}_i$ are described with the same 3 parameters and *3)* Theorem 3 shows that the image warps $\eta_{ji}$ must comply with the so-called 2D Schwarzian derivatives [27], that arise in the field of projective differential geometry.

**Theorem 2.** *If $\mathcal{M}_i$ is a 3D plane then its image embedding is $\phi_i(\mathbf{x}) = \beta_i(\mathbf{x})^{-1}(\mathbf{x} \quad 1)^\top$ with $\beta_i$ a linear function.*

*Proof.* Suppose the embedding $\mathcal{M}_i$ is a plane described by the equation $\mathbf{n}^\top\mathbf{z} + d = 0$, where $\mathbf{z} = \begin{pmatrix} z^1 & z^2 & z^3 \end{pmatrix}^\top$ and $\mathbf{n}$ is the plane's normal. From (3), the embedding is expressed with a depth function $\phi_i(\mathbf{x}) = \rho_i(\mathbf{x})\begin{pmatrix}\mathbf{x} & 1\end{pmatrix}^\top$. By combining the depth parametrization with the plane equation, we have:

$$\mathbf{n}^\top \rho_i(\mathbf{x}) \begin{pmatrix}\mathbf{x} & 1\end{pmatrix}^\top + d = 0, \quad (15)$$

from which we compute $\rho_i$ as:

$$\rho_i(\mathbf{x}) = \frac{-d}{\mathbf{n}^\top \begin{pmatrix}\mathbf{x} & 1\end{pmatrix}^\top}. \quad (16)$$

By defining $\beta_i(\mathbf{x}) = (\rho_i(\mathbf{x}))^{-1}$, $\phi_i$ is written as:

$$\phi_i(\mathbf{x}) = \beta_i(\mathbf{x})^{-1}(\mathbf{x} \quad 1)^\top. \quad (17)$$

□

Given $\mathcal{M}_i$ as a plane, its CS computed from the image embedding $\phi_i$ has a special structure that we reveal in the next corollary of Theorem 2.

**Corollary 2.** *If $\mathcal{M}_i$ is a plane then $\mathbf{\Gamma}^p_{mn}[\phi_i]$ is given by:*

$$\mathbf{\Gamma}^1_{mn}[\phi_i] = \frac{1}{\beta_i}\begin{pmatrix} -2\beta_{i1} & -\beta_{i2} \\ -\beta_{i2} & 0 \end{pmatrix} \qquad \mathbf{\Gamma}^2_{mn}[\phi_i] = \frac{1}{\beta_i}\begin{pmatrix} 0 & -\beta_{i1} \\ -\beta_{i1} & -2\beta_{i2} \end{pmatrix} \quad (18)$$

*where $\beta_{i1} = \dfrac{\partial \beta_i}{\partial x^1}$ and $\beta_{i2} = \dfrac{\partial \beta_i}{\partial x^2}$.*

*Proof.* This proof requires the manipulation of large expressions, and we thus only sketch it for the sake of readability. From the definition of $\phi_i$ in (17), we can write the Jacobian matrix of $\phi_i$ as:

$$\mathbf{J}_{\phi_i}(\mathbf{x}) = \frac{1}{\beta_i(\mathbf{x})^2}\begin{pmatrix} \beta_i(\mathbf{x}) - x^1\beta_{i1}(\mathbf{x}) & -x^1\beta_{i2}(\mathbf{x}) \\ -x^2\beta_{i1}(\mathbf{x}) & \beta_i(\mathbf{x}) - x^2\beta_{i2}(\mathbf{x}) \\ -\beta_{i1}(\mathbf{x}) & -\beta_{i2}(\mathbf{x}) \end{pmatrix}. \quad (19)$$

Note that if $\mathbf{z} = \phi_i(\mathbf{x})$, the element at the $s$th row and $k$th column of $\mathbf{J}_{\phi_i}(\mathbf{x})$ corresponds to $\frac{dz^s}{dx^k}$. Next we compute the CS by substituting (19) in (4). The metric tensor and its first derivates are then fed into (10) to obtain (18). □

**Theorem 3.** *Given that $\mathcal{M}_i$ with $i = \{1,\dots,N\}$ are infinitesimal planes, the registration warps $\eta_{ij}$ with $i,j \in \{1,\dots,N\}^2$ are point-wise solutions of the 2D Schwarzian equations.*

*Proof.* The elements of the CS for $\mathcal{M}_i$ with $i = \{1,\dots,N\}$ have the form of (18), and thus must comply with the following algebraic constraints:

$$\mathbf{\Gamma}^1_{22}[\phi_i] = \mathbf{\Gamma}^2_{11}[\phi_i] = 0 \qquad 2\mathbf{\Gamma}^2_{12}[\phi_i] = \mathbf{\Gamma}^2_{22}[\phi_i] \qquad \mathbf{\Gamma}^1_{11}[\phi_i] = 2\mathbf{\Gamma}^2_{12}[\phi_i] \quad (20)$$

where $i \in \{1,\dots,N\}$. From Corollary 1 we have $\mathbf{\Gamma}[\phi_j]^p_{mn} = \mathbf{\Gamma}[\phi_i \circ \eta_{ji}]^p_{mn}$. Now we use (11) to compute $\mathbf{\Gamma}[\phi_i \circ \eta_{ji}]^p_{nm}$ in function of $\beta_i$, $\beta_{i1}$, $\beta_{i2}$ and the derivatives of $\eta_{ji}$ up to second order. By forcing conditions in (20) in $\mathbf{\Gamma}[\phi_i \circ \eta_{ji}]$ we obtain four second order PDEs only in $\eta_{ji}$. Given that $\mathbf{y} = \eta_{ji}(\mathbf{x})$ we obtain:

$$\begin{array}{c}(\partial^2_{11}y^1)(\partial_1 y^2) - (\partial^2_{11}y^2)(\partial_1 y^1) = 0\\ (\partial^2_{22}y^1)(\partial_2 y^2) - (\partial^2_{22}y^2)(\partial_2 y^1) = 0\\ (\partial_{11}y^1)(\partial_2 y^2) - (\partial_{11}y^2)(\partial_2 y^2) +\\ 2\left((\partial_{12}y^1)(\partial_1 y^2) - (\partial_{12}y^2)(\partial_1 y^1)\right) = 0\\ (\partial_{22}y^1)(\partial_1 y^2) - (\partial_{22}y^2)(\partial_1 y^1) +\\ 2\left((\partial_{12}y^1)(\partial_2 y^2) - (\partial_{12}y^2)(\partial_2 y^1)\right) = 0\end{array} \tag{21}$$

the 2D Schwarzian equations of [22], where point-wise projective warps were investigated. □

# 4. Reconstruction Equations

## 4.1. Point-Wise Solution

We show now that local solutions to the NRSfM problem are obtained from a system of two quartics in two variables. We first select a pair of surfaces $\mathcal{M}_i$ and $\mathcal{M}_j$ and a point $\mathbf{x} = (x^1, x^2)^\top \in \mathcal{I}_i$. We evaluate $\mathbf{\Gamma}^p_{mn}[\phi_i]$ at $\mathbf{x}$, namely $\mathbf{\Gamma}^p_{mn}[\phi_i(\mathbf{x})]$, and use two unknown scalar variables, $k_1$ and $k_2$ to parametrize the CS using (18) as follows:

$$\mathbf{\Gamma}[\phi_i(\mathbf{x})]^1_{mn} = \begin{pmatrix} -2k_1 & -k_2 \\ -k_2 & 0 \end{pmatrix} \quad \mathbf{\Gamma}[\phi_i(\mathbf{x})]^2_{mn} = \begin{pmatrix} 0 & -k_1 \\ -k_1 & -2k_2 \end{pmatrix}, \tag{22}$$

where $k_1 = \frac{\beta_{i1}}{\beta_i}$ and $k_2 = \frac{\beta_{i2}}{\beta_i}$. Next we expand $\mathbf{J}_{\phi_i}$ according to (19) using $k_1$ and $k_2$:

$$\mathbf{J}_{\phi_i}(\mathbf{x}) = \frac{1}{\beta_i(\mathbf{x})} \begin{pmatrix} 1 - k_1 x^1 & -k_2 x^1 \\ -k_1 x^2 & 1 - k_2 x^2 \\ -k_1 & -k_2 \end{pmatrix} \tag{23}$$

By substitution of equation (23) into equation (4) we have:

$$\begin{array}{c}\mathbf{g}_{11}[\phi_i(\mathbf{x})] = \frac{1}{\beta_i(\mathbf{x})^2}\left(k_1^2 + (k_1 x^1 - 1)^2 + (k_1 x^2)^2\right)\\ \mathbf{g}_{12}[\phi_i(\mathbf{x})] = \frac{1}{\beta_i(\mathbf{x})^2}\left(k_1 k_2 (1 + (x^1)^2 + (x^2)^2) - k_2 x^1 - k_1 x^2\right)\\ \mathbf{g}_{22}[\phi_i(\mathbf{x})] = \frac{1}{\beta_i(\mathbf{x})^2}\left(k_2^2 + (k_2 x^1)^2 + (k_2 x^2 - 1)^2\right)\end{array} \tag{24}$$

We define $\mathbf{G}_{mn} = \beta_i(\mathbf{x})^2 \mathbf{g}_{mn}[\phi_i(\mathbf{x})]$, which only depends on $k_1$ and $k_2$. We now use $\mathbf{x} = \eta_{ji}(\mathbf{y})$ and from (11) we obtain $\mathbf{\Gamma}^p_{mn}[\phi_i \circ \eta_{ji}(\mathbf{y})]$:

$$\begin{array}{rcl}\mathbf{\Gamma}^1_{mn}[(\phi_i \circ \eta_{ji})(\mathbf{y})] & = & \begin{pmatrix} -2\bar{k}_1 & -\bar{k}_2 \\ -\bar{k}_2 & 0 \end{pmatrix}\\ \mathbf{\Gamma}^2_{mn}[(\phi_i \circ \eta_{ji})(\mathbf{y})] & = & \begin{pmatrix} 0 & -\bar{k}_1 \\ -\bar{k}_1 & -2\bar{k}_2 \end{pmatrix},\end{array} \tag{25}$$

where $\bar{k}_1$ and $\bar{k}_2$ are linear combinations of $k_1$ and $k_2$. Rewriting (24) for $\phi_j(\mathbf{y})$ we obtain $\mathbf{g}_{mn}[\phi_j(\mathbf{y})]$ in function of $\bar{k}_1$, $\bar{k}_2$ and $\beta_j(\mathbf{y})$. We then define $\overline{\mathbf{G}}_{mn} = \beta_j(\mathbf{y})^2 \mathbf{g}[\phi_i \circ \eta_{ji}(\mathbf{y})]_{mn}$. From (6) and using the definitions of $\mathbf{G}_{mn}$ and $\overline{\mathbf{G}}_{mn}$ we have the following equations:

$$\frac{1}{\beta_i(\mathbf{x})^2}\overline{\mathbf{G}}_{st} = \frac{1}{\beta_j(\mathbf{y})^2}\frac{\partial x^m}{\partial y^s}\frac{\partial x^n}{\partial y^t}\mathbf{G}_{mn}. \tag{26}$$

We cancel $\beta_i(\mathbf{x})$ and $\beta_j(\mathbf{y})$ by converting the system in (26) into the following two equations:

$$\begin{array}{c}\overline{\mathbf{G}}_{11}\left(\frac{\partial x^m}{\partial y^1}\frac{\partial x^n}{\partial y^2}\mathbf{G}_{mn}\right) - \overline{\mathbf{G}}_{12}\left(\frac{\partial x^m}{\partial y^1}\frac{\partial x^n}{\partial y^1}\mathbf{G}_{mn}\right) = 0\\ \overline{\mathbf{G}}_{11}\left(\frac{\partial x^m}{\partial y^2}\frac{\partial x^n}{\partial y^2}\mathbf{G}_{mn}\right) - \overline{\mathbf{G}}_{22}\left(\frac{\partial x^m}{\partial y^1}\frac{\partial x^n}{\partial y^1}\mathbf{G}_{mn}\right) = 0,\end{array} \tag{27}$$

which may be written in matrix form as:

$$\overline{\mathbf{G}}_{st} \propto \frac{\partial x^m}{\partial y^s}\frac{\partial x^n}{\partial y^t}\mathbf{G}_{mn}. \tag{28}$$

Equation (27) is a system of two quartics in two variables $k_1$ and $k_2$, modeling Iso-NRSfM for manifolds $\mathcal{M}_i$ and $\mathcal{M}_j$ at point $\mathbf{x} \in \mathcal{I}_i$. We denote the two equations as $\mathcal{P}_{i,j}(\mathbf{x}, k_1, k_2)$. By keeping the first index as the reference manifold, for instance $i = 1$, and obtaining the polynomials for the rest of views we obtain $2n-2$ polynomial equations in two variables $\mathcal{P}_1(\mathbf{x}, k_1, k_2) = \{\mathcal{P}_{1,j}(\mathbf{x}, k_1, k_2)\}_{j=2}^n$. The solution in $k_1$ and $k_2$ to the polynomial system $\mathcal{P}_1(\mathbf{x}, k_1, k_2)$ for a point $\mathbf{x} = \mathbf{x}_0$ allows us to reconstruct the metric tensor, the CS and the tangent plane for point $\mathbf{x}_0$ in view $\mathcal{I}_1$. Using equation (23) we can reconstruct $\mathbf{J}_{\phi_i}(\mathbf{x}_0)$ up to an unknown scale $\beta_i(\mathbf{x}_0)^{-1}$. It is not necessary to recover this scale to estimate the unitary normal, computed by taking the cross product of the two columns of $\mathbf{J}_{\phi_i}(\mathbf{x}_0)$ and normalizing.

## 4.2. Algorithm

We describe our solution to NRSfM based on the theoretical results drawn from the previous sections. The inputs of our system is a set of $N$ images of a deforming object and the outputs are the evaluated depths and normals of the deformable objects depicted in each of $N$ images. We fix a reference image such as $i = 1$, and match point correspondences between the reference image and the rest of the images. From the matched correspondences, we evaluate $\eta_{1j}$ warps using Schwarps [22] and extract a grid of points on all images. For each point $\mathbf{x}$ in the grid we can write the polynomial system described in (27). For $N$ images, we have $2N-2$ polynomials with only 2 variables $k_1$ and $k_2$ (the CS of the reference image). There are two main steps in our algorithm: *1) Find the CS of all images.* We compute a sum-of-squares of the $2N-2$ polynomials obtained from (27) and find $k_1$ and $k_2$ by minimising this sum-of-squares polynomial using [14]. By using $k_1$ and $k_2$, we can write the CS for the rest of the images using (11). *2) Evaluate depths and normals of the $N$ deformable objects.* Now that we have the CS for the grid points in all the images, the normals can be obtained by normalising the cross-product

Figure 2: Some images of the rug (top) and cat (bottom) datasets. The five rightmost images of the cat dataset are zoomed in to improve visibility.

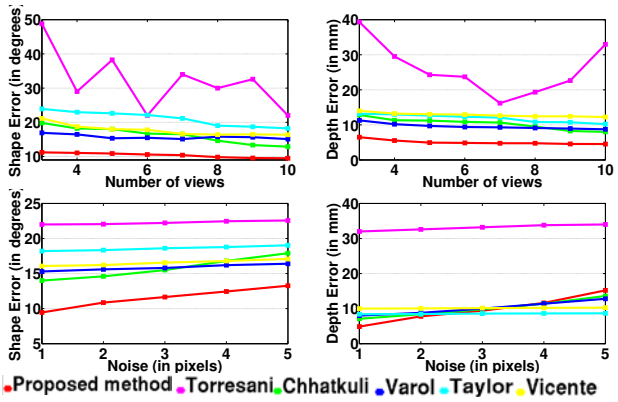


Figure 3: Synthetic data experiments. Average shape and depth errors with respect to number of views and noise.

of the two columns of the jacobian defined in (23) in terms of the CS. Then, we use the method described in [4] to recover the surfaces by integrating the normal fields.

## 5. Experimental Results

We report experiments with synthetic data and five sets of real data. We compared our method[1] with five other NRSfM methods *Chhatkuli* [4], *Varol* [1], *Taylor* [16], *Vicente* [25], *Torresani* [17], *Gotardo* [20] (only for temporal sequences). The code for these methods was obtained from the authors' websites except *Varol* which we re-implemented. We measure the shape error (mean difference between computed and ground truth normals in degrees) and depth error (mean difference between computed and ground truth 3D coordinates) to quantify the results.

**Experiments with synthetic data.** We simulated random scenes of a cylindrical surface deforming isometrically. The image size is $640p \times 480p$ and the focal length is $400p$. We tracked 400 points. We compared all methods by varying the number of views and noise in the images. The results are shown in figure 3. The results are obtained after averaging the errors over 50 trials (the default is $1p$ noise and 10 views). *On varying number of views:* Our method

[1]The code is available at https://github.com/shaifaliparashar.

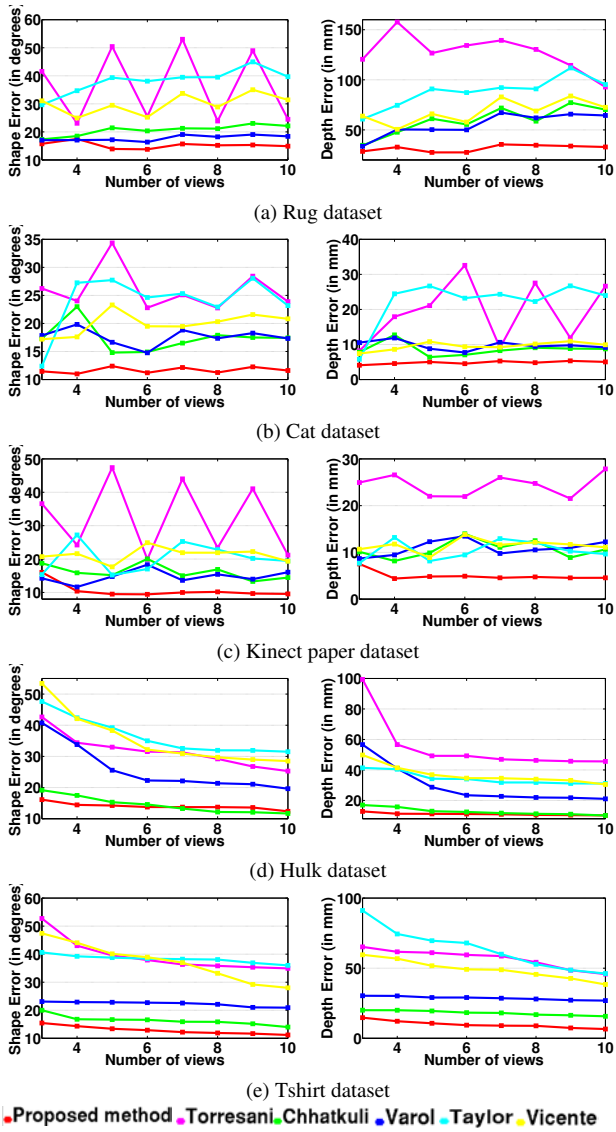


Figure 4: Real data experiments. Shape and depth errors with number of views varying from 3 to 10.

gives a very good reconstruction for 3 views which improves when more images are added. *Varol*, *Vicente* and

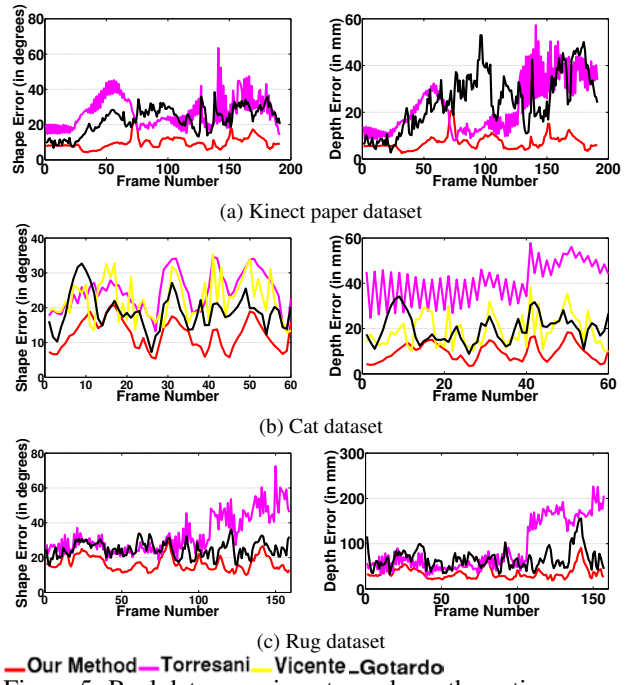


Figure 5: Real data experiments made on the entire rug, cat and kinect paper datasets.

*Chhatkuli* also perform a good reconstruction on varying number of views. *Taylor* gives decent results with 8-10 views. *Torresani* needs a video sequence or views with wide-baseline viewpoints, 10 views are not enough for reconstruction especially when they are low-baseline viewpoints. It therefore did not do well. The proposed method has consistently lower error then all others. The standard deviation of the depth error (in $mm$, for 10 images) is $2.38, 22.96, 8.78, 6.70, 6.48, 6.36$ for our method, *Torresani*, *Chhatkuli*, *Varol*, *Taylor* and *Vicente* respectively. This shows that that our method is also very consistent in terms of reconstruction. *On varying noise:* For the 10 images of the synthetic dataset, we observe that all methods change the error linearly when noise varying from 1-5 pixels is added. *Vicente* and *Taylor* show a good tolerance to noise, even though their performance is worse than other methods. Our method, *Chhatkuli* and *Varol* give higher errors with noise greater than 3 pixels. Our method gives the best performance in the 1-3 pixel noise which is what we expect in real images.

**Experiments with real data.** We conducted experiments with five datasets: the hulk and t-shirt datasets (10 different images of a paper and a cloth deformed isometrically; public dataset [4]), the kinect paper dataset (a video sequence of 191 frames and 1500 points of a paper deformed isometrically; public dataset [9]), the rug dataset (159 images of a rug deforming isometrically, captured using kinect, points obtained using [21]) and the cat dataset (60 images of a table mat deforming isometrically, captured using kinect, points obtained using [21]) (see figure 2). Our observations are

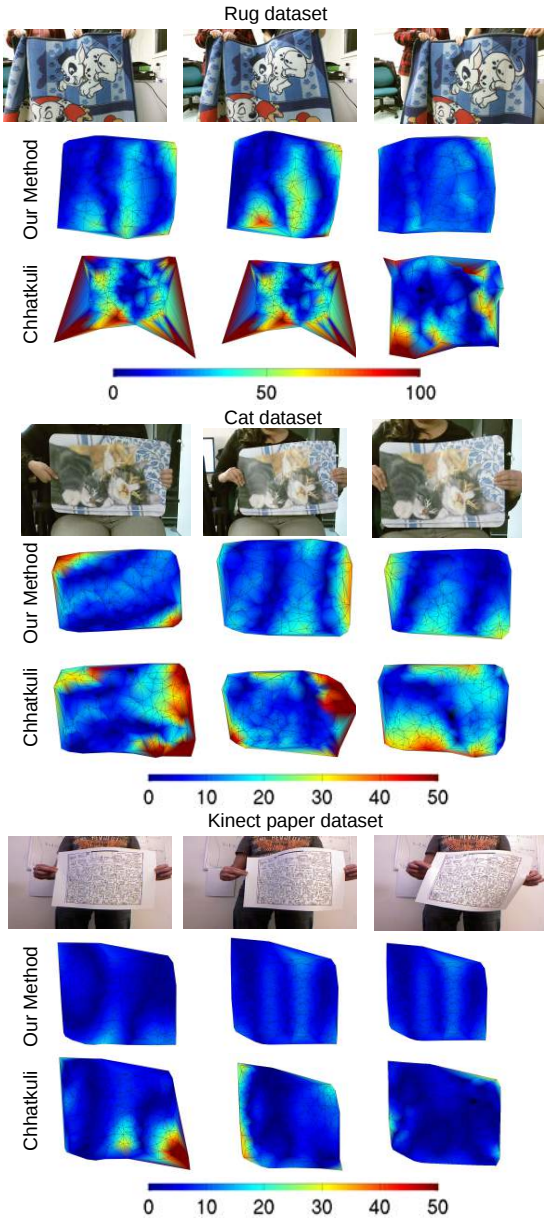


Figure 6: Reconstruction error maps for the rug, cat and kinect paper datasets. The depth errors are depicted in $mm$.

| Method | Rug | Cat | Kinect paper |
|---|---|---|---|
| Our method | 34.9 - 16.7 | 9.6 - 16.9 | 7.1 - 9.6 |
| Gotardo | 67.1 - 19.8 | 17.8 - 19.2 | 20.6 - 18.7 |
| Torresani | 92.7 - 33.3 | 40.9 - 24.7 | 22.9 - 26.9 |
| Vicente | X | 19.1 - 22.4 | X |

Table 1: Summary of methods compared on the complete sequences. Each block represents the average depth error (in $mm$) - shape error (in $^\circ$).

| % missing data | Rug | Cat | Kinect paper |
|---|---|---|---|
| 0 | 28.8 - 15.5 | 4.8 - 5.1 | 7.1 - 10.9 |
| 10 | 29.0 - 16.1 | 5.2 - 5.4 | 7.5 - 11.2 |
| 20 | 29.3 - 16.7 | 5.9 - 5.6 | 7.8 - 11.9 |
| 30 | 29.8 - 17.3 | 6.6 - 5.9 | 8.3 - 12.6 |
| 40 | 30.2 - 17.6 | 7.2 - 6.4 | 9.1 - 13.1 |
| 50 | 30.7 - 18.0 | 7.9 - 7.0 | 9.7 - 13.9 |

Table 2: Performance of our method in case of missing data. Each block represents the average depth error (in $mm$) - shape error (in $^\circ$).

summarised below. The rug, cat and kinect paper datasets are long sequences and the hulk and t-shirt are short datasets (10 images, 110 and 85 point correspondences only). We have designed two kinds of experiments: 1) Experiments where the long sequences are uniformly sampled and we compare our results with the rest of the methods (see figure 4). The results for the hulk and t-shirt datasets are averaged over 20 trials of randomly picking up images. 2) Experiments with entire sequences (see figure 5). Since our method can easily handle a large number of images, it is important to show results on large sequences. A limitation of current NRSfM methods is that they cannot handle a large number of views. Also, several NRSfM methods [4, 1] reconstruct the reference image only and are computationally expensive to recover the other shapes. *Torresani, Gotardo* reconstruct the entire imageset in one execution and therefore, we compare our method with both of them on long sequences. The cat dataset is a relatively short sequence (60 images) therefore, we added the results of *Vicente* for this dataset on the entire sequence. *Chhatkuli* and *Varol* need to compute homographies between image pairs, therefore, they grow non-linearly with the number of views. For 60 images, the execution time goes up to 45 min for a single reconstruction. Therefore, we did not compare against them even on the cat sequence. *Taylor* breaks on the cat sequence, therefore we did not include it. One must also note that *Vicente, Taylor* and *Torresani* grow with the number of views and point correspondences, therefore, they are not very efficient with a large number of views.

*Rug, cat and kinect paper datasets.* The length of the portion of the rug, table mat and paper tracked is $1m$, $35cm$ and $30cm$ respectively. Figure 4(a-c) show that our method works best amongst the compared methods for these datasets. We perform consistently better than the other methods by a significant margin in these datasets. This is quite in accordance with the results obtained for the synthetic data. *Varol, Chhatkuli* also show good results on these datasets. *Vicente* has a comparable depth error but relatively higher shape error on these datsets. This indicates that the reconstruction is more or less placed at the right place but the object is flat. *Taylor* gives bad results on the rug and cat dataset but decent results on the kinect paper dataset because it is a good dataset for orthographic methods as there is not too much perspective in the deformations. *Torresani* gives bad results because 10 views are not enough.

However, when compared against our method on the entire sequences, *Torresani* gives better results because of the higher number of views (see figure 5). *Gotardo* performs better than *Torresani* on all the sequences and gives decent results on the cat and rug dataset. Figure 6 shows the reconstruction error maps for the rug, cat and kinect paper dataset on some of the images used to compare all methods in figure 4. Table 1 summarises the performance of the compared methods on the complete sequences.

*Hulk and T-shirt datasets.* The length of the portion of the paper and cloth tracked is $24cm$ and $20cm$ respectively. Figure 4(d-e) shows that our performance is very close to *Chhatkuli* which is significantly better than other methods. *Chhatkuli* is particularly better than *Varol* on these datasets because it deals with wide-baseline data very effectively. *Torresani* gives sensible results because even though there are fewer views, the deformations are large and over wide-baseline viewpoints.

*Missing data.* Current NRSfM methods do not deal with occlusions and missing data effectively. Since our approach is local, it deals with such conditions very effectively. Table 2 shows the average depth and shape error obtained on the 3 datasets when $0 - 50\%$ data is missing from at least one of the views. The errors shown in the table are calculated only for the views which have missing data.

## 6. Conclusions

We proposed a theoretical framework for solving NRSfM locally for surfaces deforming isometrically using Riemannian geometry for manifolds for $N \geq 3$ views. Unlike other methods, the proposed method has only two variables for $N$ views. Therefore, it easily handles large numbers of views. The complexity is linear which is a substantial improvement over the current state-of-the-art methods. We tested our method on datasets with wide-baseline and short-baseline viewpoints, large and small deformations. Our results show that the proposed method consistently gives significantly better results than the state-of-the-art methods even for as few as 3 views. For future work, we will explore the possibility of extending this framework to non-isometric deformations.

**Acknowledgements.** This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE, the Spanish Ministry of Economy and Competitiveness under project SPACES-UAH (TIN2013-47630-C2-1-R) and by the University of Alcalá under project ARMIS (CCG2015/EXP-054).

## References


[1] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *CVPR*, 2009. 1, 2, 6, 8

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

[3] 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. 1

[4] A. Chhatkuli, D. Pizarro and A. Bartoli. Non-rigid shape-from-motion for isometric surfaces using infinitesimal planarity. In *BMVC*, 2014. 1, 2, 6, 7, 8

[5] 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. 1

[6] A. Agudo and F. Moreno-Noguer. Simultaneous pose and non-rigid shape with particle dynamics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2179–2187, 2015. 2

[7] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. Montiel. Sequential non-rigid structure from motion using physical priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1–1, 2015. 1, 2

[8] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in neural information processing systems*, pages 41–48, 2009. 1, 2

[9] A.Varol, M. Salzmann, P. Fua and R. Urtasun. A constrained latent variable model. In *CVPR*, 2012. 7

[10] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):567–585, 1989. 1, 2

[11] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. *International Journal of Computer Vision*, 107(2):101–122, 2014. 2

[12] R. Hartley and R. Vidal. Perspective nonrigid shape and motion recovery. In *Computer Vision–ECCV 2008*, pages 276–289. Springer, 2008. 2

[13] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000. 1

[14] D. Henrion and J.-B. Lasserre. Gloptipoly: Global optimization over polynomials with matlab and sedumi. *ACM Transactions on Mathematical Software (TOMS)*, 29(2):165–194, 2003. 5

[15] J. Fayad, A. Del Bue, L. Agapito and P. M.Q. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *BMVC*, 2009. 2

[16] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010. 1, 2, 6

[17] L. Torresani, A. Hertzmann and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008. 1, 2, 6

[18] J. Lee. *Riemannian manifolds : an introduction to curvature.* Springer, 1997. 2, 3

[19] J. Lee. *Introduction to Smooth Manifolds.* Springer, 2003. 3

[20] P.F.U. Gotardo, A.M. Martinez. Kernel non-rigid structure from motion. In *ICCV*, 2011. 1, 2, 6

[21] R. Garg, A. Roussos, L. Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314, 2013. 7

[22] R. Khan, D. Pizarro and A. Bartoli. Schwarps: Locally Projective Image Warps Based on 2D Schwarzian Derivatives. In *ECCV*, 2014. 1, 2, 5

[23] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3009–3016. IEEE, 2011. 2

[24] C. Russell, R. Yu, and L. Agapito. Video pop-up: Monocular 3d reconstruction of dynamic scenes. In *Computer Vision–ECCV 2014*, pages 583–598. Springer, 2014. 2

[25] S. Vicente and L. Agapito. Soft inextensibility constraints for template-free non-rigid reconstruction. In *ECCV*, 2012. 1, 2, 6

[26] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):931–944, 2011. 1

[27] M. Sasaki, T Yoshida. Schwarzian derivatives and uniformization. *CRM Proc Lecture Notes AMS*, (672):271–286, 2002. 4

[28] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2117–2130, 2013. 2