A Perspective on Non-Isometric Shape-from-Template

Adrien Bartoli*
Université d'Auvergne

Erol Özgür[†]
Université d'Auvergne

ABSTRACT

Shape-from-Template (SfT) uses an object's shape template and a deformation law to achieve single-image reconstruction. SfT is a fundamental tool to retexture or augment a deformable object in a monocular video. It has matured for isometric deformations, but the non-isometric case is yet largely open. This is because modeling is generally more complicated and the constraints certainly weaker. Existing algorithms use, for instance, linear elasticity, require one to provide boundary conditions represented by known deformed shape parts and need nonconvex optimization. We use a very simple and generic model to show that non-isometric SfT has a unique solution up to scale under strong perspective imaging and mild deformation curvature. Our model uses a novel type of homography interpretation that we call Perspective-Projection-Affine-Embedding. It may use boundary conditions if available and can be estimated with Linear Least Squares optimization. We provide experimental results on synthetic and real data.

1 INTRODUCTION

Shape-from-Template (SfT) aims at reconstructing an object's shape from a single image, using a template shape and a deformation law. By shape we refer to some representation of the object's 3D outer surface. SfT is applicable to objects which have a matchable appearance, which means that correspondences may be established between their template shape and the image. The template shape may correspond to any configuration of the object. It is constructed prior to using SfT by leaving the object static and 3D scanning it using for instance Shape-from-Motion (SfM) [12]. The object may then be moved and deformed: SfT will find its new shape from a single image and correspondences between the template shape and the image. As figure 1 illustrates, this makes SfT a fundamental tool to achieve Augmented Reality with a monocular camera: whether one wants to retexture or augment an object's appearance, one will have to first design the desired augmentation on the template shape, and then transfer the augmentation to the current video frame in realtime. This second step requires one to solve for correspondences between the current video frame and the template shape, and to reconstruct the shape as observed in the current video frame. This is exactly what SfT achieves, by using deformation priors, defining how the object may deform, to constrain the solution. The deformation law is a physics-based deformation prior. The isometric deformation law says deformation preserves geodesic distances. All other non-trivial deformation laws are non-isometric. Isometric SfT is an important case as it handles many objects such as paper sheets and some types of cloth. It has recently matured algorithmically [5, 16, 14] and theoretically [4, 7]. Non-isometric SfT however has not been understood as deeply. Most existing algorithms require additional cues such as shading [15] or boundary conditions [11, 13], represented by known 3D points on the object's shape. However, no theoretical results were established

---

*email:adrien.bartoli@gmail.com

[†]email:erol.ozgur@gmail.com

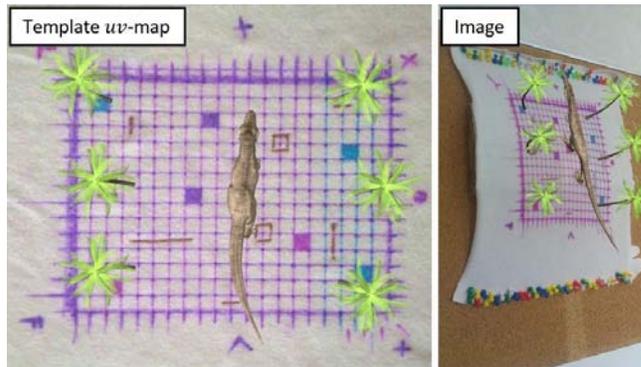regarding the number and position of these points and the problem's solvability.

Figure 1: **Shape-from-Template and Augmented Reality on a deformable surface.** This example shows a piece of fabric which was bent and substantially stretched upward, and which shrunk laterally. The user manually designed an augmentation, by adding the dinosaur and palm trees on the template, shown left. The template is a 3D deformable model, a simple plane in this example. The augmentation was then automatically transferred to the image, shown right. This required us to compute the registration between the template and the image, to know *where* to augment, and the 3D shape observed in the image, to know *how to render* the augmentation. In particular, we computed the amount of stretching, and used it to exaggeratedly rescale the dinosaur. Both image registration and shape reconstruction are outputs of SfT.

This paper contributes to the theory and algorithms of non-isometric SfT. One of the central theoretical questions in SfT is well-posedness. Because there is an infinite number of shapes which may explain the image, well-posedness questions whether the true shape is uniquely and stably recoverable. It was established that the well-posedness of isometric SfT depends on the imaging geometry: it has a unique solution for perspective imaging [4] but multiple solutions for affine imaging [7]. In non-isometric SfT, it was established that the shape cannot be resolved unless extra constraints such as multi-image temporal consistency are provided [18]. We prove a general result: the true shape is uniquely recoverable up to scale in non-isometric SfT for strongly perspective imaging and mild curvature deformation. We will make the definition of strong perspective clear. This result holds for any deformation law. It was a known result from [18], which uses a mesh representation of shape to demonstrate it. Our new demonstration is substantially different. It uses differential geometry, shows that the shape's normal is locally recoverable (which was not known from previous work) and that the reconstruction ambiguities can be raised using a curvature prior. This last point is important from a practical standpoint: previous work requires multiple images, while we achieve reconstruction from a single image. We provide five algorithms which solve non-isometric SfT in the strong-perspective-mild-curvature case. The first one follows the steps of our theoretical proof and illustrates it. The four others minimize a global cost
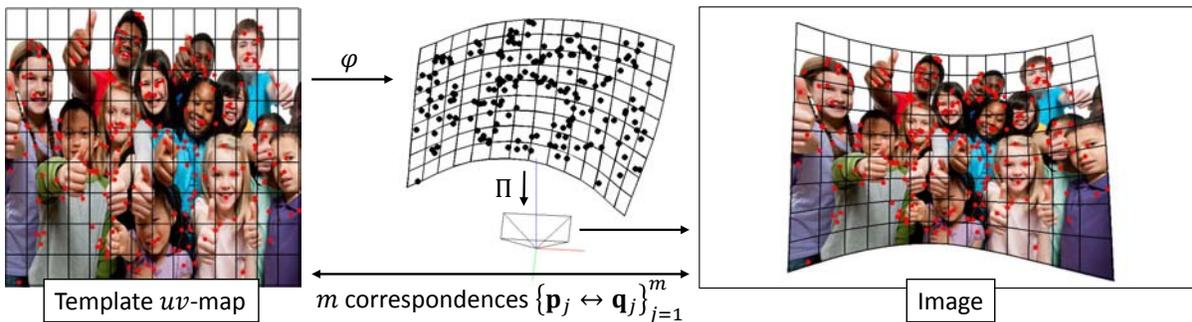
Figure 2: **Modeling Shape-from-Template.** We model the unknown shape by an embedding $\varphi \in C^2(\mathscr{T}, \mathbb{R}^3)$, with $\mathscr{T} \subset \mathbb{R}^2$ the uv-map, and the known camera by a projection function Π. In this semi-synthetic example, we used 200 SIFT keypoint correspondences, a simulated A4-size surface stretched of 25% in one axis and a 35 mm camera with a 15 mm focal length. The overlaid black grid represents ground truth.

function. Three of them are Linear Least-Squares (LLS) and thus convex and one is Nonlinear Least-Squares (NLS) and nonconvex. We provide experimental results which support the proposed theory and algorithms.

2 PREVIOUS WORK

The vast majority of existing SfT systems establish point correspondences between the template shape and the image. They then infer shape and refine the correspondences [8, 19, 20]. At the heart of an SfT system thus lies a shape inference module which reconstructs the shape from 3D to 2D point correspondences.

Non-isometric SfT is a general problem, which was instantiated with several deformation laws: conformity [4], linear elasticity [11, 13] and learnt linear shape subspace [15]. Linear elasticity was also used in NRSfM to build a linear shape subspace [2]. Except [4], these works study the problem primarily from the algorithmic and experimental viewpoints. The formulations in [11, 13] are nonconvex, involve shape globally and use boundary conditions. The formulation in [15] exploits shading to disambiguite the problem. However, it is a fundamental theoretical question in SfT to study the well-posedness for some deformation law. For instance, conformal SfT has a discrete set of solutions for perspective imaging [4]. A close work to ours is [18], which shows using a mesh-based representation that each shape point is well constrained in two degrees-of-freedom by the reprojection constraint, while its depth is very weakly constrained. For video inputs, [18] proposes to resolve this ambiguity using the shape's previous state from temporal consistency, which resembles the boundary conditions used in [11, 13]. Temporal consistency is not applicable in the single image case.

Similarly to [18], we use a generic formulation of non-isometric SfT without a specific deformation law. We show however that under some conditions, namely strong perspective and mild curvature, shape is stably recoverable from a single image without using boundary conditions and temporal consistency. More precisely, we show that the shape's normal is locally recoverable, from which shape may be recovered up to scale by curvature-regularized integration. Our results thus match those of [18], which shows that depth is globally recoverable up to scale, but not locally, and is weakly constrained. We thus bring the theoretical results that *(i)* the shape's normal is locally recoverable, even though depth is not, and *(ii)* the reconstruction ambiguities are resolved by curvature-based shape regularization, without the need to use multiple images. In practice, we use the Thin-Plate Spline (TPS) to represent functions and implement our theory algorithmically. The TPS is practical because it allows one to measure and penalize the integral curvature analytically and as a convex function of its parameters. However,

any other smooth function representation with this property, such as the tensor product of B-splines [17], could be used.

3 PROBLEM FORMULATION AND PRELIMINARIES

Setup. We formulate the problem as illustrated in figure 2. We use an embedding $\varphi \in C^2(\mathscr{T}, \mathbb{R}^3)$ as in [4] to represent the unknown shape. Here $\mathscr{T} \subset \mathbb{R}^2$ is a uv-map obtained by flattening the template shape using for instance conformal flattening [21]. We assume without loss of generality that the template shape is flat, for simplicity of the equations. We assume that m point correspondences $\mathbf{p}_j \leftrightarrow \mathbf{q}_j$, $j = 1, \ldots, m$, are given between the uv-map and the image. The camera is calibrated with intrinsics contained in matrix $\mathtt{K} \in \mathbb{R}^{3 \times 3}$. The shape is reconstructed in camera coordinates, and perspective projection is then simply $\tilde{\mathbf{q}} \propto \mathtt{K} \mathbf{Q}$ with $\tilde{\mathbf{q}}$ the homogeneous coordinates of point $\mathbf{q} \in \mathbb{R}^2$ or $\Pi(\mathbf{Q}) \overset{\text{def}}{=} \Psi(\mathtt{K}\mathbf{Q})$ with $\Psi([X\,Y\,Z]^\top) \overset{\text{def}}{=} \frac{1}{Z}[X\,Y]^\top$.

Problem formulation. We formulate non-isometric SfT with three terms:

$$\begin{cases} (\Pi \circ \varphi)(\mathbf{p}_j) = \mathbf{q}_j & j = 1, \ldots, m \\ \|D^2 \varphi\|_2 \to \min & \text{on } \mathscr{T} \\ \mathscr{W}[\varphi] = s. \end{cases} \tag{1}$$

The first term is the reprojection constraint. The second term expresses the smoothness of φ by penalizing its curvature. The third term simply fixes the shape's scale. It uses the operator $\mathscr{W}[\varphi] = \int_{\mathscr{T}} \|\varphi\|_2^2 \, d\mathbf{p}$ which returns the average square distance of the shape's points to the camera centre and sets its value to an arbitrary non-zero constant $s \in \mathbb{R}^*$. This term is important as in non-isometric SfT scale is not recoverable. Without this third term, the formulation would have $\varphi = \mathbf{0}$ as trivial solution. The third term may be replaced by boundary conditions $\mathscr{B}[\varphi] = \mathbf{0}$, if available. In an isometric formulation, the powerful geodesic distance preservation would replace the third term. At first sight, formulation (1) seems to be extremely ambiguous. The main research questions we address in this paper are *(i)* if this non-isometric formulation of SfT is solvable and *(ii)* if it may be solved numerically efficiently.

Function representation. Solving SfT numerically requires choosing a representation of the functions being involved such as the embedding φ. As already mentioned, we use the TPS representation [9]. We write $\text{TPS}(\mathscr{T})$ the space of TPS functions constructed using l control points regularly positioned in the uv-map $\mathscr{T}$, with $\text{TPS}(\mathscr{T}) \subset C^2(\mathscr{T}, \mathbb{R})$. We thus represent multivalued functions by several TPS sharing their control points. For instance,

we represent φ by three such TPS functions. Slightly abusing notation, we write as $\varphi \in \text{TPS}(\mathscr{T})$ the single and multivalued TPS functions. The TPS representation is well adapted to our problem as *(i)* it interpolates between points by minimizing the integral curvature and *(ii)* its integral curvature $\mathscr{S}[\varphi] \overset{\text{def}}{=} \int_{\mathbb{R}^2} \|D^2\varphi\|_2^2 \, d\mathbf{p}$ is a simple LLS function of its control points. We use $\mathscr{S}[\varphi]$ for the second term of formulation (1). The integral domain being $\mathbb{R}^2$ instead of $\mathscr{T}$ has no influence as curvature vanishes very quickly away from the correspondence points in $\mathscr{T}$. Importantly, $\mathscr{S}[\varphi]$ measures the out-of-plane and the in-plane curvature. In all our algorithms but one, $\mathscr{S}[\varphi]$ will be weighted by a regularization weight λ. The shape's scale $\mathscr{W}[\varphi]$ is an LLS function of the control points.

4 THEORY

4.1 Main Result on Well-Posedness

Definition 1 (Strong perspective imaging geometry). *We define the imaging geometry to be strongly perspective if the effect of perspective is observable in a local neighborhood for each shape point in the image.*

Definition 2 (Mild shape curvature). *We define the shape curvature to be mild if the shape's embedding can be locally approximated well by an affine tranformation. This means that the out-of-plane and the in-plane curvatures are both small.*

Theorem 1. *If curvature is mild and perspective strong then shape is uniquely recoverable up to scale in non-isometric SfT.*

Lemma 1. *If curvature is mild and perspective strong then the shape's normal is uniquely recoverable locally.*

Proof of theorem 1. We first show that shape is uniquely recoverable up to scale, and then that scale is unrecoverable. From lemma 1, the shape's normal is uniquely recoverable. By normal field integration, shape is then uniquely recoverable up to scale [10]. Scaling the shape does not change reprojection because of the perspective camera but scales smoothness accordingly. More precisely, we have $\mathscr{S}[k\varphi] = k^2\mathscr{S}[\varphi]$ for any $k \in \mathbb{R}^*$. Any attempt at estimating scale will thus shrink the estimated shape to the camera centre. Scale is thus unrecoverable. $\square$

4.2 A New Homography Interpretation

The interpretation of 2D homographies. An homography $\mathtt{H}$ may be estimated from image data. It can then be interpreted in a way which depends on the imaging context. For instance, the homography estimated between two images given by a purely rotating camera observing a rigid scene may be interpreted as $\mathtt{H} \propto \mathtt{KRK}^{-1}$ where $\mathtt{R}$ is the camera rotation. Plane-based pose is represented by a rigid motion $(\mathtt{R}, \mathbf{T}) \in \mathbb{SO}_3 \times \mathbb{R}^3$ which may be solved by computing the homography between the plane's model and its image, with $\mathtt{H} \propto \mathtt{K}[\mathbf{R}_1 \ \mathbf{R}_2 \ \mathbf{T}]$, $\mathbf{R}_1$ and $\mathbf{R}_2$ being the first two columns of $\mathbf{R}$. This last formula is a specialization of the well-known homography decomposition $\mathtt{H} \propto \mathtt{K}_2(d\mathtt{R} - \mathbf{t}\mathbf{n}^\top)\mathtt{K}_1^{-1}$. In non-isometric SfT however, none of the existing interpretations holds. The former is for pure rotation, the latter for plane pose, and applies locally to isometric SfT only. We thus propose a novel interpretation which applies locally to non-isometric SfT, called the Perspective-Projection-Affine-Embedding (PPAE) homography.

Affine embeddings. We show that a plane's model is related by an homography to a perspective image of its affine embedding, and that the plane's normal may be uniquely recovered from this homography. Affine embedding may be understood easily starting from the notion of rigid embedding, which is the rigid transformation mapping the plane's model to 3D in plane-based pose. A rigid embedding thus has 6 degrees-of-freedom. Affine embedding is similar, but adds anisotropic scaling and shear, and thus has 9

degrees-of-freedom. It is simply represented by $(\mathtt{A}, \mathbf{T}) \in \mathbb{R}^{3\times 2} \times \mathbb{R}^3$ and is applied as $\mathbf{p} \in \mathbb{R}^2 \to \mathtt{A}\mathbf{p} + \mathbf{T} \in \mathbb{R}^3$. The two columns of $\mathtt{A}$, $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^3$, form a basis for the embedded plane, with $\mathbf{T}$ as origin.

Perspective-Projection-Affine-Embedding homographies. We define the PPAE homography as $\mathtt{H} \propto \mathtt{K}[\mathtt{A} \ \mathbf{T}]$. It may be easily verified that this is the composition of perspective projection with an affine embedding of a plane in the camera's coordinate frame. By estimating $\mathtt{H}$ from image data, one may thus recover the affine embedding $(\mathtt{A}, \mathbf{T})$ up to scale as $[\mathtt{A} \ \mathbf{T}] \propto \mathtt{K}^{-1}\mathtt{H}$. The affine embedding has 9 degrees-of-freedom; 8 of which may be recovered from the 8 degrees-of-freedom of the homography. The plane's normal $\mathbf{N} \in \mathbb{R}^3$ can then be recovered as the normalized cross-product of the two columns of $\mathtt{A}$, as $\mathbf{N} \propto \mathbf{A}_1 \times \mathbf{A}_2$. In the absence of perspective, $\mathtt{H}$ will be an affine transformation, whose last row will be $[0 \ 0 \ a]$ with $a \neq 0$. The normal vector which may be computed in this case will necessarily be $[0 \ 0 \ 1]$, which means that the estimated plane orientation will be fronto-parallel. In other words, the observed perspective gives surface orientation.

4.3 Proof of Lemma 1

Proof. Using the assumption that the shape's curvature is mild, we may approximate the embedding locally by its first-order Taylor expansion at point $\mathbf{p} \in \mathscr{T}$ as $\varphi(\mathbf{p}') \approx \mathbf{Q} + \mathtt{J}(\mathbf{p}' - \mathbf{p})$ with $\mathbf{Q} = \varphi(\mathbf{p})$ and $\mathtt{J} = D\varphi(\mathbf{p})$. This is an affine embedding. Therefore, as we have shown, the shape normal at $\mathbf{p}$ can be estimated given the local homography $\mathtt{H}_\mathbf{p}$ induced by the shape's tangent plane at $\mathbf{p}$. Because the deformation is smooth and perspective is assumed to be strong, the homography $\mathtt{H}_\mathbf{p}$ can be estimated from image data. One possible way is to use the optic flow field between the *uv*-map and the image in the vicinity of $\mathbf{p}$. According to [18], the optic flow field may be estimated for any point given sufficiently many correspondences. We represent it by an image warp $\eta \in C^2(\mathscr{T}, \mathbb{R}^2)$ which may be computed by smoothly densifying the correspondences. We define the image warp η as a function mapping points from the *uv*-map to the image. The warp provides 2 constraints at zeroth-order on the homography as:

$$\eta(\mathbf{p}) \ = \ \Psi(\mathtt{H}_\mathbf{p}\tilde{\mathbf{p}}). \tag{2}$$

Differentiating this constraint to first- and second-orders provides 4 and 6 additional constraints respectively, from which the 8 degrees-of-freedom of homography $\mathtt{H}_\mathbf{p}$ can be estimated. $\square$

5 A NUMERICAL ALGORITHM FOLLOWING THE PROOFS

We give an algorithm which implements a proof of concept for our theory by following the ideas used in the proofs of theorem 1 and lemma 1. We name this algorithm TI for non-isometric Tangent-plane-Integration. TI computes local surface information which is then integrated to recover shape. Rather than computing the normal field, it uses the more informative tangent-plane field. It is illustrated in figure 3.

Differentiation of the optic flow field at the correspondences. The goal of this step is to estimate the flow field's first and second derivatives at the correspondences. This may be achieved in many ways. For instance, one may exploit the image intensity around each correspondence to fit a local warp. In TI, we fit a warp $\eta \in \text{TPS}(\mathscr{T})$ with automatic estimation of the regularization weight [3]. We have $\eta(\mathbf{p}) = [\eta_x(\mathbf{p}) \ \eta_y(\mathbf{p})]^\top$ and $\mathbf{p} = [u \, v]^\top$. The sought derivatives are then obtained in closed-form. We denote the 4 first-order derivatives as $\nabla_{\{u,v\}}\eta_{\{x,y\}}(\mathbf{p})$ and the 6 second-order derivatives as $\nabla_{\{uu,vv,uv\}}\eta_{\{x,y\}}(\mathbf{p})$.

Tangent-plane field estimation. This step estimates the tangent-plane field at the correspondences. This requires one to first estimate a homography field and then extract the tangent-planes using the PPAE interpretation. An homography $\mathtt{H}_j$ is estimated around each correspondence $\mathbf{p}_j \leftrightarrow \mathbf{q}_j$, using the optic flow field's
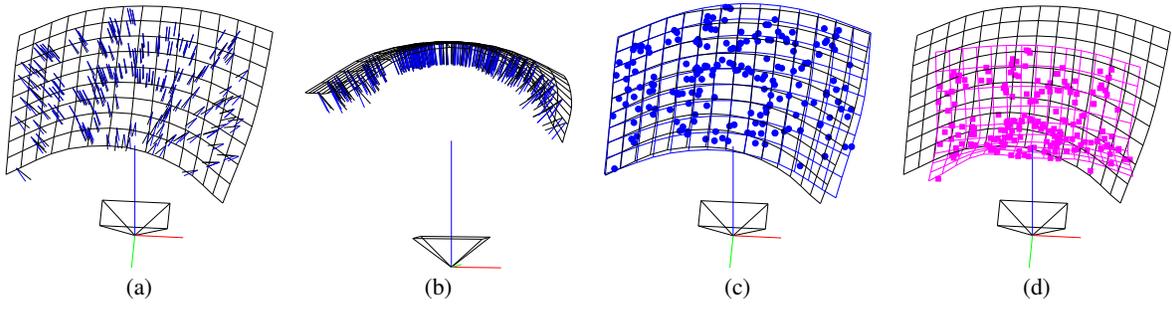
Figure 3: **Results of TI.** (a) and (b) show the reconstructed normal field and (c) the reconstructed points and surface in blue. The normal error is 6.7 degrees and the position error is 44.6 mm. (d) shows the result of an isometric algorithm [16]. Ground truth is in black.
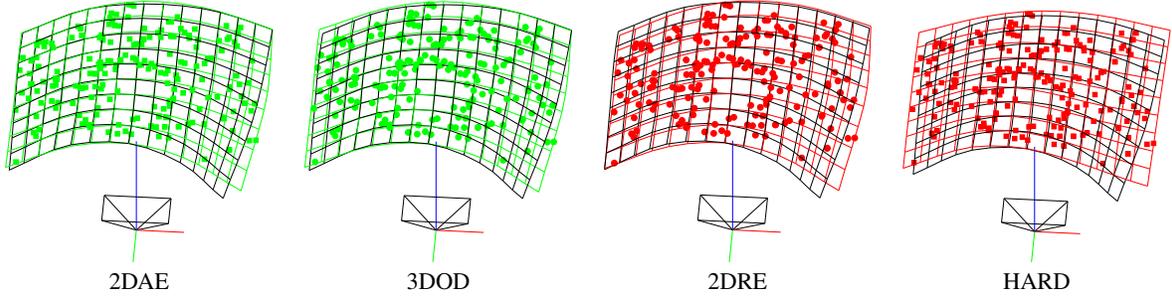

Figure 4: **Results of the proposed direct global numerical algorithms.** The normal errors are 6.5, 6.0, 7.8 and 10.7 degrees. The position errors are 42.7, 40.0, 60.4 and 63.4 mm.

zeroth-, first- and second-order derivatives, as in [6]. The main difference is that [6] derived nonlinear constraints, while we derive linear constraints, which we can thus solve easily. Defining vect as row-wise vectorization, we start by rewriting the reprojection constraint (2) as:

$$\zeta_0(\mathbf{p}_j)\text{vect}(\mathtt{H}_j) = \mathbf{0}_{2\times1} \quad \text{with} \quad \zeta_0(\mathbf{p}) \stackrel{\text{def}}{=} \begin{bmatrix} -\tilde{\mathbf{p}}^\top & \mathbf{0}_{1\times3} & \eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top \\ \mathbf{0}_{1\times3} & -\tilde{\mathbf{p}}^\top & \eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top \end{bmatrix},$$

We call this equation the *Algebraic Transfer Constraint (ATC)* of zeroth-order. This is because when minimized in the least-squares sense this equation gives the algebraic distance used in the Direct Linear Transform (DLT) [12, §4.1]. However, *(i)* we have only one point, *(ii)* instead of a fixed target point, we have a warp function which provides extra constraints by differentiation. The first- and second-order ATCs are then obtained by simply differentiating ζ_0.

Defining $\tilde{\mathtt{I}} \stackrel{\text{def}}{=} [\mathtt{I}_{2\times2}\,\mathbf{0}_{2\times1}]$ as a truncated 3×3 identity matrix and its two rows as $\mathbf{e}_1^\top$ and $\mathbf{e}_2^\top$ we have:

$$\zeta_1(\mathbf{p}) \stackrel{\text{def}}{=} \begin{bmatrix} \tilde{\mathtt{I}} & \mathbf{0}_{2\times3} & \begin{matrix} \nabla_u\eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top + \eta_x(\mathbf{p})\mathbf{e}_1^\top \\ \nabla_v\eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top + \eta_x(\mathbf{p})\mathbf{e}_2^\top \end{matrix} \\ \mathbf{0}_{2\times3} & \tilde{\mathtt{I}} & \begin{matrix} \nabla_u\eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top + \eta_y(\mathbf{p})\mathbf{e}_1^\top \\ \nabla_v\eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top + \eta_y(\mathbf{p})\mathbf{e}_2^\top \end{matrix} \end{bmatrix}$$

$$\zeta_2(\mathbf{p}) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0}_{6\times6} & \begin{matrix} \nabla_{uu}\eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top + 2\nabla_u\eta_x(\mathbf{p})\mathbf{e}_1^\top \\ \nabla_{uu}\eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top + 2\nabla_u\eta_y(\mathbf{p})\mathbf{e}_1^\top \\ \nabla_{vv}\eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top + 2\nabla_v\eta_x(\mathbf{p})\mathbf{e}_2^\top \\ \nabla_{vv}\eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top + 2\nabla_v\eta_y(\mathbf{p})\mathbf{e}_2^\top \\ \nabla_{uv}\eta_x(\mathbf{p})\tilde{\mathbf{p}}^\top + \nabla_v\eta_x(\mathbf{p})\mathbf{e}_1^\top + \nabla_u\eta_x(\mathbf{p})\mathbf{e}_2^\top \\ \nabla_{uv}\eta_y(\mathbf{p})\tilde{\mathbf{p}}^\top + \nabla_v\eta_y(\mathbf{p})\mathbf{e}_1^\top + \nabla_u\eta_y(\mathbf{p})\mathbf{e}_2^\top \end{matrix} \end{bmatrix}.$$

We estimate the homography $\mathtt{H}_j$ by finding the least singular vector of a (12×9) matrix using an SVD [12, A5.3] to solve the following

norm-constrained homogeneous LLS problem:

$$\min_{\mathtt{H}_j\in\mathbb{R}^{3\times3}} \sum_{i=1}^{3} \|\zeta_i(\mathbf{p}_j)\text{vect}(\mathtt{H}_j)\|_2^2 \quad \text{s.t.} \quad \|\mathtt{H}_j\|_2 = 1. \quad (3)$$

Using the PPAE homography interpretation, we finally estimate a basis for the tangent plane as the two columns of $\mathtt{G}_j \stackrel{\text{def}}{=} \mathtt{K}^{-1}\mathtt{H}_j\tilde{\mathtt{I}}^\top$, $j = 1,\ldots,m$.

Shape-from-Tangent-Plane. This is the integration step, which estimates the embedding φ representing the shape. The basis for the tangent plane is given by $D\varphi$, and we thus have the constraints $D\varphi(\mathbf{p}_j) \propto \mathtt{G}_j$, $j = 1,\ldots,m$. We translate these constraints into a minimizable LLS penalty by introducing an unknown vector $\mathbf{a} = [a_1\cdots a_m]^\top \in \mathbb{R}^m$ such that the squared L_2 norm of $D\varphi(\mathbf{p}_j) - a_j\mathtt{G}_j$ can be minimized. We also incorporate the constraints from problem (1). The bending energy $\mathscr{S}[\varphi]$ is already expressed as a minimizable LLS penalty and the scale constraint $\mathscr{W}[\varphi] = s$ is used directly as a quadratic norm constraint. We use the algebraic error for the reprojection constraint, which we write as the squared L_2 norm of $\tilde{\mathtt{I}}\left[\tilde{\mathbf{q}}_j\right]_\times \mathtt{K}\varphi(\mathbf{p}_j)$, where $[\cdot]_\times$ is the skew-symmetric cross-product matrix. This leads to the following homogeneous LLS problem with an $(8m+3l)\times(m+3l)$ design matrix representing:

$$\min_{\substack{\varphi\in\text{TPS}(\mathscr{T})\\\mathbf{a}\in\mathbb{R}^m}} \sum_{j=1}^{m} \left(\|D\varphi(\mathbf{p}_j) - a_j\mathtt{G}_j\|_2^2 + \|\tilde{\mathtt{I}}\left[\tilde{\mathbf{q}}_j\right]_\times \mathtt{K}\varphi(\mathbf{p}_j)\|_2^2 \right) + \lambda\mathscr{S}[\varphi]$$

$$\text{s.t.} \quad \mathscr{W}[\varphi] = s \text{ or } \mathscr{B}[\varphi] = \mathbf{0}.$$

6 DIRECT GLOBAL NUMERICAL ALGORITHMS

Instantiating and solving problem (1) by directly searching for an embedding satisfying all constraints is a much simpler type of algorithm than TI. The rationale is that all the information derived
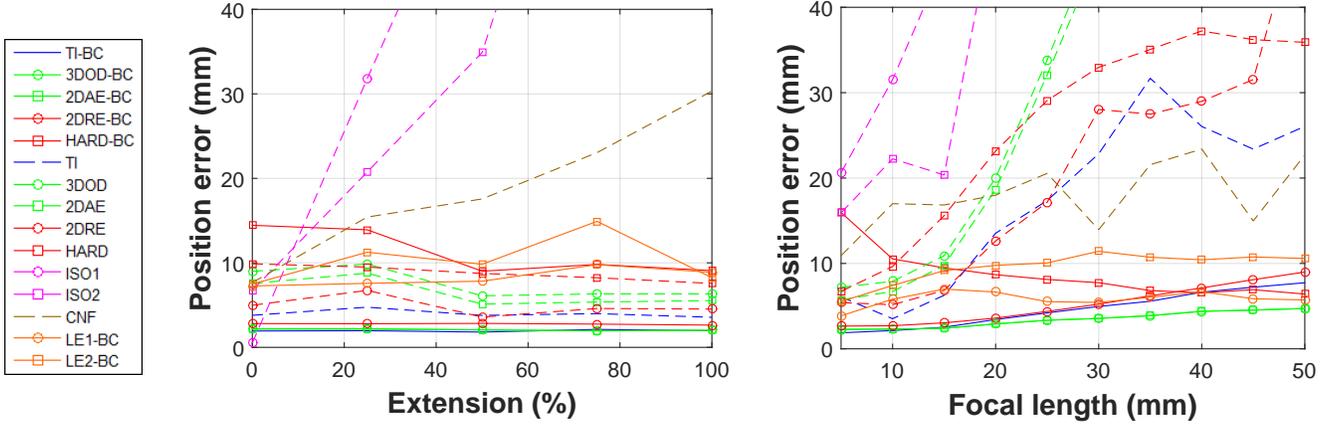
Figure 5: **Results on simulated data.** The x-axis is the percentage of simulated surface extension (left) and the focal length (right).

through the steps of TI such as the tangent-plane field is contained in the correspondence data and derived from the initial constraints of problem (1). We give four possible ways to instantiate and solve problem (1) numerically. They all use the surface's bending energy $\mathscr{S}[\varphi]$ and scale constraint $\mathscr{W}[\varphi] = s$ or boundary conditions $\mathscr{B}[\varphi] = \mathbf{0}$ and differ in how they instantiate the reprojection constraint. They are illustrated in figure 4.

2DAE: minimizing the 2D Algebraic Error. Our first algorithm instantiates the reprojection constraint with the square algebraic error. It is important as it is strictly similar to the integration step of TI, but without the penalty on the tangent-plane field. We arrive at an homogeneous LLS formulation with a $(2m+3l) \times 3l$ design matrix representing:

$$\min_{\varphi \in \text{TPS}(\mathscr{T})} \sum_{j=1}^{m} \left\| \tilde{\mathtt{I}} \left[\tilde{\mathbf{q}}_j \right]_\times \mathtt{K}\varphi(\mathbf{p}_j) \right\|_2^2 + \lambda \mathscr{S}[\varphi] \text{ s.t. } \mathscr{W}[\varphi] = s \text{ or } \mathscr{B}[\varphi] = \mathbf{0}.$$

3DOD: minimizing the 3D Orthogonal Distance. We instantiate the reprojection constraint with the square distance between the predicted shape point $\varphi(\mathbf{p}_j)$ and the sight-line of point $\mathbf{q}_j$. We arrive at an homogeneous LLS formulation with a $(3m+3l) \times 3l$ design matrix representing:

$$\min_{\varphi \in \text{TPS}(\mathscr{T})} \sum_{j=1}^{m} \left\| \left(\mathtt{I} - \frac{\mathtt{K}\tilde{\mathbf{q}}_j \tilde{\mathbf{q}}_j^\top \mathtt{K}^\top}{\tilde{\mathbf{q}}_j^\top \mathtt{K}^\top \mathtt{K}\tilde{\mathbf{q}}_j} \right) \varphi(\mathbf{p}_j) \right\|_2^2 + \lambda \mathscr{S}[\varphi]$$
$$\text{s.t. } \mathscr{W}[\varphi] = s \text{ or } \mathscr{B}[\varphi] = \mathbf{0}.$$

2DRE: minimizing the 2D Reprojection Error. We instantiate the reprojection constraint with the square distance between the predicted image point $\Pi(\varphi(\mathbf{p}_j))$ and the corresponding point $\mathbf{q}_j$. We arrive at an affine NLS formulation with $2m+3l$ equations and $3l$ unknowns, which we solve with Levenberg-Marquardt [12, A6] initialized by the result from 3DOD:

$$\min_{\varphi \in \text{TPS}(\mathscr{T})} \sum_{j=1}^{m} \left\| \Pi(\varphi(\mathbf{p}_j)) - \mathbf{q}_j \right\|_2^2 + \lambda \mathscr{S}[\varphi] \text{ s.t. } \mathscr{W}[\varphi] = s \text{ or } \mathscr{B}[\varphi] = \mathbf{0}.$$

HARD: exact reprojection. In contrast to the three above algorithms which instantiate the reprojection constraint by a penalty, HARD enforces it exactly. This is a sensible algorithm to try as we use weak priors and so data must be used primarily to constrain the

solution. We arrive at an homogeneous LLS formulation:

$$\min_{\varphi \in \text{TPS}(\mathscr{T})} \sum_{j=1}^{m} \mathscr{S}[\varphi]$$
$$\text{s.t. } \left\{ (\Pi \circ \varphi)(\mathbf{p}_j) = \mathbf{q}_j \right\}_{j=1}^{m} \text{ and } \mathscr{W}[\varphi] = s \text{ or } \mathscr{B}[\varphi] = \mathbf{0}.$$

Interestingly, this algorithm has not regularization weight as it only minimizes the bending energy. We solve it using the constrained SVD algorithm [12, A5.4] with a $3m \times 3m$ design matrix and a $2m \times 3m$ constraint matrix.

7 EXPERIMENTAL EVALUATION

Compared algorithms and measured errors. We compared the five proposed algorithms with two isometric algorithms (ISO1 [5] and ISO2 [16]), a conformal algorithm (CNF [4]) and two Linear Elastic algorithms (LE1-BC [13] and LE2-BC [11]). BC is for Boundary Conditions. We give the results of the proposed algorithm with and without boundary conditions (named as 3DOD-BC and 3DOD, for instance). All algorithms have tunable parameters; we chose them by trial and error. We measured the position and normal errors in mm and degrees respectively, as the averages of the point distance and angular normal difference to ground truth.

Simulated data. We simulated a deforming and extending surface from which we synthesized images with a setup similar to figure 2 and adding a 1 pixel magnitude noise to image points. We averaged results over 10 subsets of 100 points, and used 20 points as boundary conditions. The results are shown in figure 5. Isometric algorithms do not handle extension beyond a few percents, neither does the conformal algorithm, as the extension is non-isotropic. The linear elastic and all proposed algorithms cope extremely well with extension. The proposed algorithms without boundary conditions perform well for focal length shorter than about 20 mm. The linear elastic and all proposed algorithms with boundary conditions degrade very slightly as focal length increases, except HARD-BC which slightly improves.

Real data. We drew a regular grid on a piece of extensible fabric. We then pinned its upper and lower parts down to a fixed cork board, and inserted an object between the fabric and the board to create extension and curvature, as shown in figure 6. The extension is 33% on the y-axis and causes shrinking of 20% on the x-axis. We took 25 photos with a 35 mm camera and a 18 mm focal length, from which we used dense SfM [1] to compute ground truth. For boundary conditions we used the 17 points forming the template
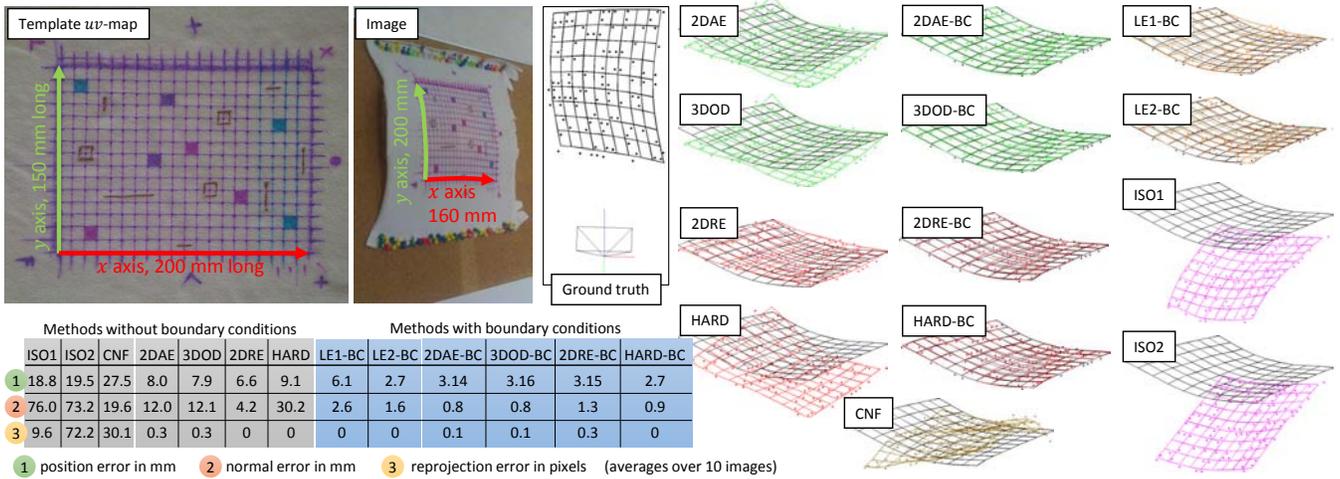
Template uv-map — y axis, 150 mm long — x axis, 200 mm long

Image — y axis, 200 mm long — x axis, 160 mm

Ground truth

2DAE · 2DAE-BC · LE1-BC · 3DOD · 3DOD-BC · LE2-BC · 2DRE · 2DRE-BC · ISO1 · HARD · HARD-BC · ISO2 · CNF

	Methods without boundary conditions							Methods with boundary conditions					
	ISO1	ISO2	CNF	2DAE	3DOD	2DRE	HARD	LE1-BC	LE2-BC	2DAE-BC	3DOD-BC	2DRE-BC	HARD-BC
1	18.8	19.5	27.5	8.0	7.9	6.6	9.1	6.1	2.7	3.14	3.16	3.15	2.7
2	76.0	73.2	19.6	12.0	12.1	4.2	30.2	2.6	1.6	0.8	0.8	1.3	0.9
3	9.6	72.2	30.1	0.3	0.3	0	0	0	0	0.1	0.1	0.3	0

1 position error in mm   2 normal error in mm   3 reprojection error in pixels   (averages over 10 images)

Figure 6: **Results on real data.** The ground-truth is shown in black in the right plots.

shape's convex hull. We ran all algorithms on 10 photos and report average errors and show example reconstructions in figure 6. This matches the observations from simulated data. Isometric algorithms fail. The proposed algorithms without boundary conditions are slightly less accurate than with boundary conditions. LE1 is less accurate than the proposed algorithms with boundary conditions, while LE2 is equivalent for position error. However, the proposed algorithms with boundary conditions have lower normal errors.

8 CONCLUSION

We have studied a generic and simple formulation of non-isometric SfT. Thanks to PPAE, a novel type of homography interpretation, we have shown that with strong perspective and mild curvature, the formulation is solvable. This extends the local analytic solutions of isometric SfT [4]. Intuitively, an isometric deformation is locally rigid, and the local image scale thus gives depth while shear and anisotropy give orientation. To solve non-isometric SfT with PPAE, we assumed that deformation is locally affine, in other words that curvature is mild. The local image scale, shear and anisotropy are then related to deformation, and not to pose, but we showed that orientation is then uniquely recoverable from local perspective.

We have given five numerical algorithms. Three of them solve the formulation globally by optimizing a convex cost, and may use boundary conditions if available. In contrast, existing non-isometric SfT algorithms are restricted to specific models, require tuning and nonconvex optimization and have not been shown to be solvable. In our experiments, our algorithms outperformed existing ones in terms of position and normal errors. Our future work will focus on *(i)* adapting the proposed methods to lenses such as the fisheye producing a very short effective focal length, *(ii)* understanding the relationship between the proposed model and linear elasticity and *(iii)* developing the method's usage to achieve the automatic augmentation of non-isometric surfaces, by extending our results from figure 1.

REFERENCES

[1] Agisoft. Photoscan, April 2015. v1.1.5.

[2] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. *CVPR*, 2014.

[3] A. Bartoli. Maximizing the predictivity of smooth deformable image warps through cross-validation. *JMIV*, 31(2-3):133–145, July 2008.

[4] A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins, and D. Pizarro. Shape-from-template. *PAMI*, 37(10):2099–2118, October 2015.

[5] F. Brunet, A. Bartoli, and R. Hartley. Monocular template-based 3D surface reconstruction: Convex inextensible and nonconvex isometric methods. *CVIU*, 125:138–154, August 2014.

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

[7] A. Chhatkuli, D. Pizarro, and A. Bartoli. Stable template-based isometric 3D reconstruction in all imaging conditions by linear least-squares. *CVPR*, 2014.

[8] T. Collins and A. Bartoli. Using isometry to classify correct/incorrect 3D-2D correspondences. *ECCV*, 2014.

[9] J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10:5–12, 1976.

[10] J.-D. Durou and F. Courteille. Integration of a normal field without boundary condition. *Workshop at ICCV*, 2007.

[11] N. Haouchine, J. Dequidt, M.-O. Berger, and S. Cotin. Single view augmentation of 3D elastic objects. *ISMAR*, 2014.

[12] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. Second Edition.

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

[14] F. Moreno-Noguer and P. Fua. Stochastic exploration of ambiguities for nonrigid shape recovery. *PAMI*, 35(2):463–475, February 2013.

[15] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua. Capturing 3D stretchable surfaces from single images in closed form. *CVPR*, 2009.

[16] J. Ostlund, A. Varol, D. Ngo, and P. Fua. Laplacian meshes for monocular 3D shape recovery. *ECCV*, 2012.

[17] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registation using Free-Form Deformations: Application to breast MR images. *TMI*, 18(8):712–721, August 1999.

[18] M. Salzmann, V. Lepetit, and P. Fua. Deformable surface tracking ambiguities. *CVPR*, 2007.

[19] J. Sanchez-Riera, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence, and non-rigid shape. *CVPR*, 2010.

[20] A. Shaji, A. Varol, L. Torresani, and P. Fua. Simultaneous point matching and 3D deformable surface reconstruction. *CVPR*, 2010.

[21] A. Sheffer, B. Lévy, M. Mogilnitsky, and A. Bogomyakov. ABF++: Fast and robust angle based flattening. SIGGRAPH, 2005.