A Stable Analytical Framework for Isometric Shape-from-Template by Surface Integration

Ajad Chhatkuli[a], Daniel Pizarro[b,a], Adrien Bartoli[a], and Toby Collins[a]

[a]ISIT - CNRS/Université d'Auvergne, Clermont-Ferrand, France

[b]GEINTRA, Universidad de Alcalá, Alcalá de Henares, Spain

Abstract

Shape-from-Template (SfT) reconstructs the shape of a deforming surface from a single image, a 3D template and a deformation prior. For isometric deformations, this is a well-posed problem. However, previous methods which require no initialization break down when the perspective effects are small, which happens when the object is small or viewed from larger distances. That is, they do not handle all projection geometries. We propose stable SfT methods that accurately reconstruct the 3D shape for all projection geometries. We follow the existing approach of using first-order differential constraints and obtain local analytical solutions for depth and the first-order quantities: the depth-gradient or the surface normal. Previous methods use the depth solution directly to obtain the 3D shape. We prove that the depth solution is unstable when the projection geometry tends to affine, while the solution for the first-order quantities remain stable for all projection geometries. We therefore propose to solve SfT by first estimating the first-order quantities (either depth-gradient or surface normal) and integrating them to obtain shape. We validate our approach with extensive synthetic and real-world experiments and obtain significantly more accurate results compared to previous initialization-free methods. Our approach does not require any optimization, which makes it very fast.

CONTENTS

I. Introduction

Context: The 3D reconstruction of deforming objects with a monocular camera has been a subject of great interest. Although the theory of rigid reconstuction by Structure-from-Motion (SfM) has matured [Hartley and Zisserman, 2004], non-rigid reconstruction remains an open problem. Monocular 3D reconstruction of deforming scenes has potential applications in a wide range of scenarios such as medical endoscopic augmented reality [Maier-Hein et al., 2014]. The difficulty in monocular non-rigid reconstruction stems from the fact that it is a severely under-constrained problem without additional priors on the surface deformation or shape space. We study the isometric deformation model, where the surface deforms such that the geodesic distance remains unchanged. The isometric model applies to a wide variety of real objects. Using the isometric model, two main non-rigid reconstruction problems were studied in the literature. In Non-Rigid Shape-from-Motion (NRSfM) [Bregler et al., 2000; Chhatkuli et al., 2014b; Collins and Bartoli, 2010; Del Bue, 2008; Taylor et al., 2010; Torresani et al., 2008; Vicente and Agapito, 2012], the 3D shape is obtained from multiple images. In Shape-from-Template (SfT) [Bartoli and Collins, 2013; Bartoli et al., 2015; Ngo et al., 2015; Perriollat et al., 2011; Salzmann and Fua, 2011], the 3D shape is obtained from a single image and a template of the object. The template is a textured 3D model of the object in a known reference position. We here study the shape inference step of SfT, assuming that registration between the template and image is solved using, for example, point correspondences [Collins and Bartoli, 2014b; Pilet et al., 2008; Pizarro and Bartoli, 2012]. For SfT, the isometric model provides locally as many constraints as the rigid planar model. A scene's projection geometry can range from strongly perspective to virtually affine. In the latter case, this happens when the object is very small or viewed from a large distance. We define a *stable method* as one that gives an accurate 3D reconstruction for all projection geometries. We propose two stable methods that require no initialization and are fast.

Several successful SfT methods that have been proposed are based on the isometric constraint. In essence, these methods differ from each other by the way isometry is imposed and how the final constraints are optimized. [Bartoli and Collins, 2013; Bartoli et al., 2015] describe isometric SfT as a Partial Differential Equation (PDE) system, giving local analytical solutions and a proof that isometric SfT is well-posed. [Brunet et al., 2014] finds the shape by minimizing a statistically optimal cost while imposing isometric constraints. [Collins and Bartoli, 2015] propose a real-time tracking-based approach for the perspective cameras. A different class of methods [Ngo et al., 2015; Perriollat et al., 2011; Salzmann and Fua, 2011] relaxes isometry with inextensibility. Using the so-called *Maximum Depth Heuristic* (MDH), [Perriollat et al., 2011; Salzmann and Fua, 2011] choose the shape that maximizes depth under the inextensibility constraint. Inextensibility means the distance between the neighboring points remain inferior to their geodesic distance in the template. Recently [Ngo et al., 2015] modified the approach of [Salzmann and Fua, 2011] by enforcing Laplacian smoothness in the solution. The methods discussed form a solid foundation for the SfT problem. However they lack accuracy and applicability in many scenarios. We claim that the following qualities are desired of an SfT method: *a)* it should be robust to noise and outliers in correspondences, *b)* it should be accurate even with a low number of matched feature points, *c)* it should work with camera focal lengths that are very small to very large (stability) and *d)* it should be fast and analytical with no requirement for an initialization. The correspondence outliers mentioned in the property

a) can be either tackled at the registration step or the reconstruction step. There are several successful methods that remove outliers during registration [Collins and Bartoli, 2014b; Pilet et al., 2008; Pizarro and Bartoli, 2012] while small noise in correspondences has to be dealt with in the reconstruction step. For that reason, we focus our further discussions of property *a)* in the context of noise in correspondences for the reconstruction step. The inextensibility-based methods [Ngo et al., 2015; Perriollat et al., 2011; Salzmann and Fua, 2011] fail to capture the property in *c)* and do not entirely satisfy *d)*, while the analytical solutions in [Bartoli and Collins, 2013; Bartoli et al., 2015] fail to capture the requirements in *c)*. Similarly the statistically optimal cost minimization in [Brunet et al., 2014; Collins and Bartoli, 2015] does not satisfy *d)*. Thus there is a clear need for a method that satisfies all four criteria.

Local analytical methods: The local analytical solutions for SfT were first given in [Bartoli et al., 2015], where a set of non-holonomic solutions was obtained for a PDE system with a change of variable. The solutions of a PDE system are called the non-holonomic solutions when they are obtained by treating any quantities and their derivatives as separate independent unknowns. Specifically, the non-holonomic solutions give the radial component of the depth and its gradient in the spherical coordinate system. Despite the fact that the local analytical depth solution is unique, we show that *it is not well-constrained when the camera projection tends to affine*. On the other hand, we prove that the depth-gradient solution is always stable. This is a significant discovery because most methods for SfT rely on computing a solution for the depth [Bartoli and Collins, 2013; Bartoli et al., 2015; Collins et al., 2014; Ngo et al., 2015; Salzmann and Fua, 2011]. We further give an alternative approach to isometric SfT, which gives an equivalent but different PDE system, inspired from the work of [Collins and Bartoli, 2014a] on plane-based pose estimation. In [Collins and Bartoli, 2014a] the pose of a rigid plane is estimated using the non-holonomic solutions of a PDE system. While [Collins and Bartoli, 2014a] does not give solutions for a deforming surface, we adapt the PDE system for SfT for both planar and non-planar templates. The resulting PDE system is equivalent to the PDE system proposed in [Bartoli et al., 2015] but its non-holonomic solutions describe different quantities. In particular its second non-holonomic solution is used differently in the subsequent steps of reconstruction that yields slightly different results. It also gives a more intuitive description of SfT and its solutions in terms of rigid transforms of tangent planes on the surface, and surface normals.

Stable solutions and stable methods: We define a *stable solution* as a non-holonomic solution of the SfT PDE system that remains well-constrained regardless of the projection geometry involved (i.e., perspective or affine). We define a *stable method* as one which solves SfT accurately for all projection geometries. We propose two stable methods that satisfy properties *a)* to *d)* described previously. We achieve this by using the stable solutions based on two first-order quantities: the depth-gradient and surface normal. In our first method, which we refer to as the *stable type-I method*, we specifically use the *radial depth-gradient solution*. We obtain this from the PDE system of [Bartoli et al., 2015]. However, the radial depth gradient is only known up to sign. We resolve the sign from the depth solution, and then integrate the quantity over the surface to obtain the radial depth values. Because this is from the integration, the values are up to a global scale factor. We compute the scale factor from the average of the depth solution. Finally we obtain the depth values by a change of variable. Our second method, the *stable type-II method* is similar, however we use non-holonomic solution for the surface normal. Like the radial depth gradient, the surface normal also has a

two-fold ambiguity. We resolve this again by using the depth solution. Finally we integrate the normals, then resolve the reconstruction's scale using the average of the depth solution. In practice, the results of the two stable methods differ slightly due to the influence of noise on the subsequent steps. We find the stable type-II method to be slightly superior to the stable type-I. Both of the proposed methods rival the accuracy of statistically optimal approaches [Brunet et al., 2014; Collins and Bartoli, 2015], are stable under any projection geometry and require no initialization. They are extension of our previous work [Chhatkuli et al., 2014a], where we explored the instability of depth solution in the differential isometric SfT formulation and proposed an algorithm to obtain a better 3D reconstruction using the stable type-I method.

Paper layout: The paper is organized as follows. We present previous work on SfT in §II. We then formally introduce our SfT problem statement and give the general PDEs describing it in §III. We present the type-I solutions, analyze their stability and give the stable type-I method in §IV. We introduce the type-II solutions, analyze their stability and give the stable type-II method in §V. We give details of the experiments and discuss the results in §VI, and finally provide our conclusions in §VII.

Notation: We use bold for vectors and matrices and italics for scalars. We use Greek letters for functions (*e.g.* ψ). We use the operator J_ψ to write the function giving the Jacobian matrix of ψ. We use $\mathrm{diag}(s_1, \ldots, s_n)$ to define a diagonal matrix from the scalars $s_1, \ldots, s_n$. Given a matrix $\mathbf{M}$ we write $\lambda_i(\mathbf{M})$ and $\mathbf{v}_i(\mathbf{M})$ for its ith eigenvalue and eigenvector. We assume the eigenvalues to be in descending order: $\lambda_i \geq \lambda_j$ with $i < j$. We denote an identity matrix of size $n \times n$ as $\mathbf{I}_n$. For a matrix $\mathbf{M}$ of size $m \times n$ we denote its sub-matrix made of the first i rows and j columns as $[\mathbf{M}]_{ij}$. We write the jth column of a matrix $\mathbf{M}$ as $[\mathbf{M}]_j$.

II. Previous Work

Work in non-rigid 3D reconstruction problems started with the template-less NRSfM scenarios. The work of [Bregler et al., 2000] is one of the earliest examples of NRSfM methods and is based on the factorization approach for an orthographic camera. The SfT methods on the other hand, were introduced later but they matured quickly and are mostly based on physics-based models of surfaces, mainly the isometric model. Isometry can be formulated as a differential constraint on the deformation function [Brunet et al., 2014] or approximated by putting bounds on the distances between neighboring points [Perriollat et al., 2011; Salzmann and Fua, 2011]. In order to organize the discussion on previous works in isometric SfT, we classify the methods based on their final constraints and the way they are optimized: *i)* zeroth-order methods based on inextensibility [Ngo et al., 2015; Salzmann and Fua, 2011], *ii)* statistically optimal cost refinement [Brunet et al., 2014; Collins and Bartoli, 2015] and *iii)* analytical solutions from quadratic PDEs [Bartoli and Collins, 2013; Bartoli et al., 2015]. The MDH-based methods in *i)* [Perriollat et al., 2011; Salzmann and Fua, 2011] solve the SfT problem by maximizing depth while putting an upper bound on the distance between neighboring points. [Perriollat et al., 2011] proposed the MDH with inextensibility as a relaxation of the original isometric constraint. [Salzmann and Fua, 2011] reformulated the MDH as a convex problem and applied Second-Order Cone Programming (SOCP) to compute depth. They also made use of a learned space of deformations using a linear local model for each

small patch on the mesh. Such an approach limits the method to the use of surfaces where learned models for patches already exist. [Ngo et al., 2015] proposed a modified approach, where the method uses the inextensibility constraints but does not maximize the depth explicitly. It uses a mesh Laplacian to parametrize the surface with a smaller set of control points. This effectively reduces the size of the final inequalities that are to be optimized by splitting the problem into two optimizations. The Laplacian parametrization removes the learning mechanism of [Salzmann and Fua, 2011] and puts a prior on the space of deformations using a method that existed in computer graphics [Sorkine et al., 2004; Sumner and Popović, 2004]. Such priors are well suited for most applications because they do not make statistical assumptions about the material properties of surfaces. However, the method requires solving two different optimization problems, one of them being non-convex. The solution requires an iterative approach and the computation time does not scale up linearly with increase in the mesh size. The zeroth-order methods are non-analytical and consequently some of their theoretical aspects are still not well understood. Nevertheless, it is clear that inextensibility does not constrain depth strongly in near-affine conditions as for the perspective conditions.

The method [Brunet et al., 2014] in *ii)* optimizes a statistically optimal cost, that includes the differential isometric constraint, 3D back-projection constraints and smoothness. On the one hand it is non-convex and relies on iterative local optimization such as Levenberg-Marquardt. Such a refinement involves the use of an initialization and requires a much higher computation time. Apart from that, the optimization requires relative weighting of the three constraints with two parameters that need to be precisely tuned to get optimal results. On the other hand, it carries the advantage of having a holonomic solution of depth, where the relationship between depth and its Jacobian is taken into account. This also means that it does not suffer from the depth instability in near-affine conditions as is the case for other methods. The method of [Collins and Bartoli, 2015] is more general than [Brunet et al., 2014] and works in real-time. It handles arbitrary surface meshes and solves the registration and shape inference problem together using dense point matches. However, it also requires an initial solution, which is obtained by tracking the object frame-to-frame.

The analytical methods in *iii)* [Bartoli and Collins, 2013; Bartoli et al., 2015] use a flattened template-to-image registration warp and the surface parametrization function derivatives to directly compute the surface's depth analytically. At each surface point, the depth is obtained as the non-holonomic solutions to a PDE system. It uses the zeroth and first-order information like *ii)* but the solutions are obtained analytically. This requires computing the gradient of the local image warp. Methods in *iii)* discard the depth-gradient solution and keep only the depth solution. The local analytical nature of the solutions means that the methods are very fast and can be parallelized efficiently as the solution for each point is found independently. As an advantage of being analytical, methods in *iii)* also form a powerful tool to analyze the effect of different projection geometries on the recovered shape.

III. Differential Geometric and PDE-based Modeling

We follow the problem modeling of [Bartoli et al., 2015] as depicted in figure 1. We start with the 3D template $\mathcal{T} \subset \mathbb{R}^3$. The flat template is a 2D space $\Omega \subset \mathbb{R}^2$ obtained from $\mathcal{T}$, which is parametrized with $\Delta \in C^1(\Omega, \mathbb{R}^3)$. Consequently the flattening function is Δ^{-1}. In practice it is obtained from a conformal flattening of a texture-mapped

mesh or often simply by taking an image of the 3D object. The 3D template $\mathcal{T}$ is transformed by an isometric deformation $\psi \in C^1(\mathcal{T}, \mathbb{R}^3)$. The deformed surface $\mathcal{S} \subset \mathbb{R}^3$ is projected by a known camera projection function Π onto an image $\mathcal{I} \subset \mathbb{R}^2$. We define $\mathcal{S}$ in the camera's coordinate frame. We denote the registration between Ω and $\mathcal{I}$ as $\eta \in C^1(\Omega, \mathbb{R}^2)$. We parametrize the deformed surface $\mathcal{S}$ by an unknown embedding function $\varphi \in C^1(\Omega, \mathbb{R}^3)$.

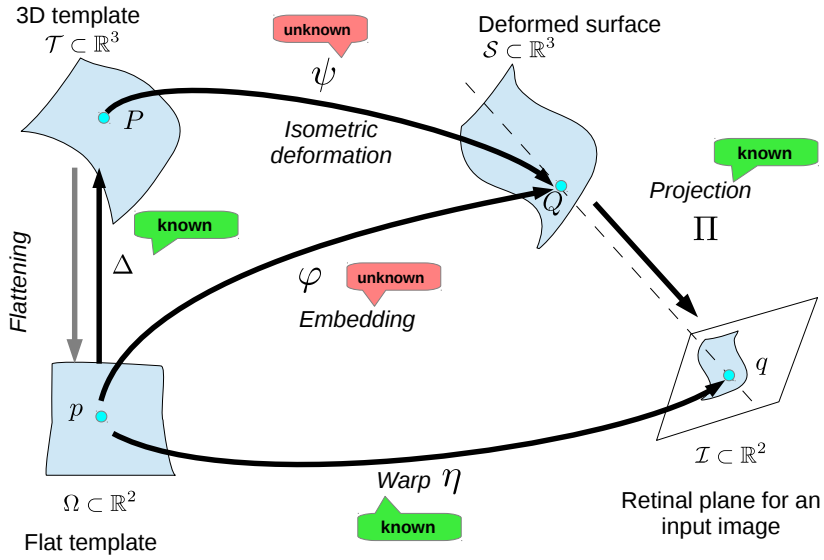


Fig. 1: Differential geometric modeling of Shape-from-Template.

Our goal is to solve the SfT problem, represented by ψ. In practice we work with the embedding φ. This is equivalent since $\varphi = \psi \circ \Delta$. We obtain φ from the known functions Δ, η and Π, and the fact that the surface deforms isometrically. Below we describe the differential constraints and then define the SfT problem with a set of PDEs.

A. *Differential Constraints*

We divide the constraints on φ into the deformation constraint and the reprojection constraint. The deformation constraint imposes the isometric prior while the reprojection constraint is analogous to a data term, which ensures that the reprojection of the surface matches the input image. This information is related to the camera geometry and imaging via the known projection function Π. We give details for both constraints as presented in [Bartoli and Collins, 2013; Bartoli et al., 2015] and additionally generalize them for different camera models with which we obtain a generalized equation in §IV.

1) Deformation constraint: We start with the equation for the embedding $\varphi = \psi \circ \Delta$; its differentiation leads to:

$$\mathsf{J}_\varphi = (\mathsf{J}_\psi \circ \Delta)\, \mathsf{J}_\Delta. \tag{1}$$

Pre-multiplying equation (1) by its transpose gives us:

$$\mathsf{J}_\varphi^\top \mathsf{J}_\varphi = \mathsf{J}_\Delta^\top (\mathsf{J}_\psi \circ \Delta)^\top (\mathsf{J}_\psi \circ \Delta)\, \mathsf{J}_\Delta. \tag{2}$$

Isometric deformations preserve geodesic distances and for such deformations we have:

$$(\mathsf{J}_\psi \circ \Delta)^\top (\mathsf{J}_\psi \circ \Delta) = \mathbf{I}_3, \tag{3}$$

which simply states that the metric tensor on the surface remains unchanged with an isometry described by ψ. Substituting equation (3) in equation (2) gives:

$$\mathsf{J}_\varphi^\top \mathsf{J}_\varphi = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta. \tag{4}$$

Equation (4) states that the first fundamental form is preserved in an isometry. Thus it remains the same for the surface embedding φ and the template parametrization Δ.

2) Reprojection constraint: The reprojection constraint is obtained with the reprojection equation:

$$\eta = \Pi \circ \varphi. \tag{5}$$

Equation (5) enforces consistency between the warp η and the projection of the embedding in the image. Without loss of generality we assume that the world coordinate frame is the camera's and we denote as $f > 0$ the camera's focal length. We then use the reprojection constraint to express the embedding $\varphi = [\varphi_x \ \varphi_y \ \varphi_z]^\top$ with the depth function $\varphi_z \in C^1(\Omega, \mathbb{R})$. We consider two possible camera models: the perspective camera and the infinitesimal weak-perspective camera.

The perspective camera: With perspective projection Π^{P} we have:

$$\eta = \Pi^{\mathrm{P}} \circ \varphi = \begin{bmatrix} f\frac{\varphi_x}{\varphi_z} & f\frac{\varphi_y}{\varphi_z} \end{bmatrix}^\top. \tag{6}$$

Using equation (6) the embedding φ may be parametrized by the depth function φ_z and the template-to-image warp function in homogeneous coordinates $\tilde{\eta}^\top = [\eta^\top \ 1]$ as:

$$\varphi = \Phi^{\mathrm{P}} \tilde{\eta} \quad \text{with} \quad \Phi^{\mathrm{P}} = \operatorname{diag}\left(\frac{\varphi_z}{f}, \frac{\varphi_z}{f}, \varphi_z\right). \tag{7}$$

The infinitesimal weak-perspective camera: This camera model was proposed to simplify the PDEs by approximating the gradient [Bartoli et al., 2013]. It is based on the weak-perspective model, which approximates the perspective camera [Hartley and Zisserman, 2004]. It first projects the scene orthographically onto a fronto-parallel plane placed at the scene's average depth and then scales it. The infinitesimal weak-perspective camera instantiates a weak-perspective camera at each point. This gives the same projection as the perspective camera but simplifies the expression for the depth-gradient. It is non-analytical. The infinitesimal weak-perspective projection Π^{WP} yields:

$$\eta = \Pi^{\mathrm{WP}} \circ \varphi = \begin{bmatrix} f\frac{\varphi_x}{\zeta} & f\frac{\varphi_y}{\zeta} \end{bmatrix}^\top. \tag{8}$$

In this model ζ represents the depth. It is different at each point and given by φ_z, while preserving the property that $\mathsf{J}_\zeta = \mathbf{0}_{1\times 2}$. The back-projection equation with the infinitesimal weak-perspective model is:

$$\varphi = \Phi^{\text{WP}}\tilde{\eta} \quad \text{with} \quad \Phi^{\text{WP}} = \text{diag}\left(\frac{\zeta}{f}, \frac{\zeta}{f}, \varphi_z\right). \tag{9}$$

Unified camera model: We give a unified model for a general camera by rewriting the reprojection constraint with the back-projection matrix as:

$$\varphi = \Phi\tilde{\eta} \quad \text{with} \quad \Phi \in \{\Phi^{\text{P}}, \Phi^{\text{WP}}\}. \tag{10}$$

The partial derivatives of the back-projection matrix Φ is:

$$\mathbf{M} = \frac{\partial \Phi}{\partial \varphi_z} = \begin{cases} \mathbf{M}^{\text{P}} = \text{diag}(\frac{1}{f}, \frac{1}{f}, 1) & \text{perspective} \\ \\ \mathbf{M}^{\text{WP}} = \text{diag}(0, 0, 1) & \text{infinitesimal weak-perspective.} \end{cases} \tag{11}$$

B. General PDE

The constraints described in §III-A were first used in [Bartoli et al., 2015] to analytically solve isometric SfT and to show that it is a well-posed problem. Here, we give the following generalized PDE system to describe SfT:

$$\text{Find } \varphi \text{ s.t.} \begin{cases} \mathsf{J}_\varphi^\top \mathsf{J}_\varphi = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta \\ \\ \Pi \circ \varphi = \eta. \end{cases} \tag{12}$$

In the following sections we reformulate system (12) and show that it intrinsically has three non-holonomic unknowns and three equations. The holonomic solutions for system (12), which enforces the differential dependency between the solutions may not exist in practice in the presence of noise. We study the space of solutions of system (12) using two different PDE formulations, which result in type-I solutions and type-II solutions. The type-I solutions provide the radial component of depth and its gradient. The type-II solutions give the depth and the surface normal. The two formulations provide the same direct depth solution, however as we will show, the second non-holonomic solutions for type-I and type-II solutions are used differently in the subsequent steps that produce accurate but slightly different reconstructions.

IV. Type-I Solutions Stability and Type-I Stable Method

The well-posedness of isometric SfT defined by the PDE system (12) was proved in [Bartoli et al., 2015] where the local non-holonomic solutions for the PDE system were derived. We first describe the analytical method to obtain the non-holonomic solutions as given by [Bartoli and Collins, 2013; Bartoli et al., 2015]. In order to achieve that, they reformulate system (12) so that it consists of three non-holonomic unknowns: the depth and its gradient with respect to the flat template. Our contribution here is that we generalize the equations for two different camera models and derive solutions for both models.

A. Type-I Solutions

We consider the embedding parametrized by the depth function φ_z as shown in equation (10). The problem of SfT can be viewed as that of finding the depth function φ_z so that the deformation constraints (4) are met. We derive a non-linear PDE system that holds for both perspective and infinitesimal weak-perspective projection (10) and deformation constraints (4). We first differentiate equations (7) and (9) and use the expressions defined in equation (11) to get J_{φ}:

$$\mathsf{J}_{\varphi} = \mathbf{M}\tilde{\eta}\mathsf{J}_{\varphi_z} + \Phi\mathsf{J}_{\tilde{\eta}}, \tag{13}$$

To verify equation (13) one can expand equation (7) or (9) before differentiating them. Similarly, following through the matrix multiplications in equation (13) we reach the same result.

We introduce equation (13) in the deformation constraint (4) to obtain the following non-linear PDE system:

$$\mathsf{J}_{\varphi_z}^{\top}\tilde{\eta}^{\top}\mathbf{M}^2\tilde{\eta}\mathsf{J}_{\varphi_z} + \mathsf{J}_{\varphi_z}^{\top}\tilde{\eta}^{\top}\mathbf{M}\Phi\mathsf{J}_{\tilde{\eta}} + \mathsf{J}_{\tilde{\eta}}^{\top}\Phi\mathbf{M}\tilde{\eta}\mathsf{J}_{\varphi_z} + \mathsf{J}_{\tilde{\eta}}^{\top}\Phi^2\mathsf{J}_{\tilde{\eta}} \quad = \mathsf{J}_{\Delta}^{\top}\mathsf{J}_{\Delta}. \tag{14}$$

System (14) models SfT in terms of φ_z and J_{φ_z} for perspective and infinitesimal weak-perspective projections. Assuming J_{φ_z} and φ_z are independent variables, we obtain the non-holonomic solutions of system (14) analytically. We denote them as $\bar{\varphi}_z \in C^1(\Omega, \mathbb{R})$ and $\bar{\kappa} \in C^0(\Omega, \mathbb{R}^2)$. Despite the fact that system (14) admits exact solutions for both $\bar{\varphi}_z$ and $\bar{\kappa}$, they are not generally consistent since $\mathsf{J}_{\bar{\varphi}_z} \neq \bar{\kappa}$. With errors in η, system (14) is in fact an overdetermined PDE system with no general (*i.e.* holonomic) solutions.

1) Perspective camera: The PDE system (14) is specialized to perspective projection by choosing Φ^{P} from equation (7) and $\mathbf{M}^{\mathrm{P}}$ from equation (11):

$$\left(1 + \frac{\eta^{\top}\eta}{f^2}\right)\mathsf{J}_{\varphi_z}^{\top}\mathsf{J}_{\varphi_z} + \frac{\varphi_z}{f^2}(\mathsf{J}_{\varphi_z}^{\top}\eta^{\top}\mathsf{J}_{\eta} + \mathsf{J}_{\eta}^{\top}\eta\mathsf{J}_{\varphi_z}) + \frac{\varphi_z^2}{f^2}\mathsf{J}_{\eta}^{\top}\mathsf{J}_{\eta} = \mathsf{J}_{\Delta}^{\top}\mathsf{J}_{\Delta}.$$

We simplify this system by changing variables with:

$$\alpha = \varphi_z\nu \quad \text{and} \quad \nu = \sqrt{1 + \frac{\eta^{\top}\eta}{f^2}}, \tag{15}$$

giving $\mathsf{J}_{\alpha} = \nu\mathsf{J}_{\varphi_z} + \frac{\varphi_z}{\nu f^2}\eta^{\top}\mathsf{J}_{\eta}$. This leads to an equivalent but simpler PDE system in α and J_{α}:

$$\mathsf{J}_{\alpha}^{\top}\mathsf{J}_{\alpha} + \alpha^2\gamma = \mathsf{J}_{\Delta}^{\top}\mathsf{J}_{\Delta}, \tag{16}$$

where:

$$\gamma = \frac{1}{\nu^2 f^2}\left(\mathsf{J}_{\eta}^{\top}\mathsf{J}_{\eta} - \frac{1}{\nu^2 f^4}\mathsf{J}_{\eta}^{\top}\eta\eta^{\top}\mathsf{J}_{\eta}\right). \tag{17}$$

In order to obtain the non-holonomic solutions we subsitute J_{α} by an independent vector function $\beta \in C^0(\Omega, \mathbb{R}^2)$ in equation (16). This gives us:

$$\beta^{\top}\beta + \alpha^2\gamma = \mathsf{J}_{\Delta}^{\top}\mathsf{J}_{\Delta}. \tag{18}$$

Following [Bartoli and Collins, 2013; Bartoli et al., 2015] we can always find a single algebraic solution of system (18). We denote the non-holonomic solutions of α and β as $\bar{\alpha}$ and $\bar{\beta}$. To obtain them we first modify equation (18) as:

$$\beta^\top \beta \gamma^{-1} = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta \gamma^{-1} - \alpha^2 \mathbf{I}_2. \tag{19}$$

The left-hand side of equation (19) clearly has rank 1 and therefore the second eigenvalue of the right hand side is 0 as well. With that we obtain:

$$\bar{\alpha} = \sqrt{\lambda_2 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \gamma^{-1} \right)}. \tag{20}$$

Substituting the solution for α in equation (18) the non-holonomic solution $\bar{\beta}$ is obtained as:

$$\bar{\beta} = \pm \sqrt{\lambda_1(\Upsilon)} \mathbf{v}_1(\Upsilon), \tag{21}$$

where:

$$\Upsilon = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \lambda_2 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \gamma^{-1} \right) \gamma. \tag{22}$$

We may recover $\bar{\varphi}_z$ from equation (20) followed by the change of variable (15). Instead of $\bar{\kappa}$, we recover $\bar{\beta}$ from equation (21). The solutions $\bar{\alpha}$ and $\bar{\beta}$ are the non-holonomic solutions of the perspective type-I PDE system (18) obtained algebraically and thus they exist in all practical circumstances. In our stable type-I method we give an alternative for obtaining the depth at each point using $\bar{\beta}$.

2) Infinitesimal weak-perspective camera: The PDE for the infinitesimal weak-perspective camera is found by choosing Φ^{WP} from equation (9) and $\mathbf{M}^{\text{WP}}$ from equation (11):

$$\mathsf{J}_{\varphi_z}^\top \mathsf{J}_{\varphi_z} + \frac{\zeta^2}{f^2} \mathsf{J}_\eta^\top \mathsf{J}_\eta = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta, \tag{23}$$

where we set $\zeta = \varphi_z$ in the infinitesimal weak-perspective model as ζ gives the 'average' depth at a differential level.

System (23) has exactly the same structure as system (16), the simplified PDE system for perspective projection. We directly obtain the non-holonomic solutions of the system $\bar{\varphi}_z$ and $\bar{\kappa}$, without any change of variable. To obtain $\bar{\varphi}_z$ and $\bar{\kappa}$ in the infinitesimal weak-perspective model we first assign $\gamma^{\text{WP}} = f^{-2} \mathsf{J}_\eta^\top \mathsf{J}_\eta$. This transforms equation (23) into the following:

$$\kappa^\top \kappa + \varphi_z^2 \gamma^{\text{WP}} = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta. \tag{24}$$

Noting the similarity of equation (24) with equation (18), we give the non-holonomic solutions for equation (24) as:

$$\bar{\varphi}_z = \sqrt{\lambda_2 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \left(\gamma^{\text{WP}} \right)^{-1} \right)} \quad \text{and} \quad \bar{\kappa} = \pm \sqrt{\lambda_1(\Upsilon^{\text{WP}})} \mathbf{v}_1(\Upsilon^{\text{WP}}), \tag{25}$$

where:

$$\Upsilon^{\text{WP}} = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \lambda_2 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \left(\gamma^{\text{WP}} \right)^{-1} \right) \gamma^{\text{WP}}.$$

3) Obtaining the embedding: [Bartoli and Collins, 2013; Bartoli et al., 2015] use $\bar{\varphi}_z$ directly to get the embedding

φ through equation (7) or (9), neglecting the information contained in $\bar{\beta}$ and consequently $\bar{\kappa}$. At first glance this direct-depth method seems to be sensible as $\bar{\beta}$ is known only up to sign and requires integration to recover depth. We show in §IV-B however, that the depth solution $\bar{\varphi}_z$ is not well-constrained, unlike the depth-gradient. We give a method of obtaining a better 3D reconstruction using $\bar{\beta}$ in §IV-C.

B. Stability

We prove two important results regarding the stability of the type-I non-holonomic solutions of PDEs (16) and (23). We list them as propositions below.

Proposition 1. *The non-holonomic solution for depth $\bar{\varphi}_z$ is weakly constrained when the projection geometry tends to affine.*

Proposition 2. *The non-holonomic solution for the depth-gradient $\bar{\kappa}$ is well-constrained in all projection geometries.*

Figure 3 gives a general diagram showing the effect of different projection geometries on SfT. We prove these results for the perspective and infinitesimal weak-perspective cameras.

We define a projection function Π_s on a 3D point $\mathbf{Q} = [Q_x \; Q_y \; Q_z]^\top$, depending on a parameter s that allows us to continuously select the amount of perspective:

$$\Pi_s(\mathbf{Q}) = \frac{(s+1)\,f}{Q_z + sf}\begin{bmatrix} Q_x & Q_y \end{bmatrix}^\top. \tag{26}$$

We obtain a perspective projection with the focal length f when $s = 0$, and an orthographic projection[1] when $s \to \infty$:

$$\lim_{s\to\infty} \Pi_s(\mathbf{Q}) = \begin{bmatrix} Q_x & Q_y \end{bmatrix}^\top. \tag{27}$$

The infinitesimal weak-perspective approximation of Π_s is:

$$\Pi_s^{\text{WP}}(\mathbf{Q}) = \frac{(s+1)\,f}{\zeta + sf}\begin{bmatrix} Q_x & Q_y \end{bmatrix}^\top. \tag{28}$$

Proof of propositions 1 and 2 for the perspective camera. We first substitute the projection model Π_s into the PDE system (14) by simply redefining the back-projection matrix Φ for perspective projection as:

$$\Phi_s = \operatorname{diag}\left(\frac{\varphi_z + sf}{(s+1)f}, \frac{\varphi_z + sf}{(s+1)f}, \varphi_z + sf\right). \tag{29}$$

Introducing Φ_s in the type-I PDE system (14) we obtain:

$$\left(1 + \frac{\eta^\top\eta}{((s+1)f)^2}\right)\mathsf{J}_{\varphi_z}^\top\mathsf{J}_{\varphi_z} + \frac{\varphi_z + sf}{((s+1)f)^2}(\mathsf{J}_{\varphi_z}^\top\eta^\top\mathsf{J}_\eta) + \frac{\varphi_z + sf}{((s+1)f)^2}(\mathsf{J}_\eta^\top\eta\mathsf{J}_{\varphi_z}) + \frac{(\varphi_z + sf)^2}{((s+1)f)^2}\mathsf{J}_\eta^\top\mathsf{J}_\eta = \mathsf{J}_\Delta^\top\mathsf{J}_\Delta. \tag{30}$$

[1]The orthographic projection and affine projection are equivalent up to an affine image transform. Thus the discussion is still valid for any affine projection although described for an orthographic projection.

We first prove proposition 1. By taking the limit $s \to \infty$ in equation (30) we find the following system:

$$\mathsf{J}_{\varphi_z}^\top \mathsf{J}_{\varphi_z} + \mathsf{J}_\eta^\top \mathsf{J}_\eta = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta, \tag{31}$$

which represents the general PDE system for affine projection [Pizarro et al., 2013]. In equation (31) the depth variable φ_z vanishes, which means that in affine projection depth is not any more constrained by the perspective model. This can be also proved in the space of solutions after the change of variable with equation (15). $\bar{\alpha}$ depends on the eigenvalues of matrix $\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \gamma^{-1}$. When s is a large number, the solution of $\bar{\alpha}$ in equation (20) is ill-conditioned. We write γ from equation (17) as a function of s:

$$\gamma_s = \frac{1}{\nu_s^2 \left((s+1)f\right)^2} \left(\mathsf{J}_\eta^\top \mathsf{J}_\eta - \frac{1}{\nu_s^2 \left((s+1)f\right)^4} \mathsf{J}_\eta^\top \eta\eta^\top \mathsf{J}_\eta \right), \tag{32}$$

with:

$$\nu_s = \sqrt{1 + \frac{\eta^\top \eta}{((s+1)f)^2}}. \tag{33}$$

Taking the limit of equation (32) we find $\lim_{s\to\infty} \gamma_s = \mathbf{0}_{2\times 2}$. $\bar{\alpha}$ is then computed from a matrix whose elements tend to infinity.

As for proposition 2, although φ_z has already vanished from equation (31), the solution to its gradient can still be recovered by applying the rank-1 constraint to equation (31). Thus $\bar{\kappa}$ is simply given by:

$$\bar{\kappa} = \pm\lambda_1 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \mathsf{J}_\eta^\top \mathsf{J}_\eta \right) \mathbf{v}_1 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \mathsf{J}_\eta^\top \mathsf{J}_\eta \right). \tag{34}$$

which means that depth-gradient is equally well constrained with affine projection. □

Proof of propositions 1 and 2 for the infinitesimal weak-perspective camera. By taking the infinitesimal weak-perspective approximation of the projection model Π_s^{WP} and plugging it into equation (23) after setting $\zeta = \varphi_z$ we reach the following system:

$$\mathsf{J}_{\varphi_z}^\top \mathsf{J}_{\varphi_z} + \left(\frac{\varphi_z + sf}{(s+1)f} \right)^2 \mathsf{J}_\eta^\top \mathsf{J}_\eta = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta. \tag{35}$$

Again by taking the limit $s \to \infty$ on both sides of equation (35), we reach system (31) of affine projection. If there is no perspective effect and the camera is affine we cannot compute the average depth of the scene as it vanishes from the equations. The proof of proposition 1 follows in the same way as in the perspective camera by using $\gamma_s = \frac{1}{((s+1)f)^2} \mathsf{J}_\eta^\top \mathsf{J}_\eta$. For proposition 2 the solution of $\bar{\kappa}$ in equation (35) is identical to that for the perspective model when $s \to \infty$. □

C. Stable Type-I Methods

In §IV-A and IV-B we revisited the non-holonomic type-I solutions and proved that the depth solution was unstable. However obtaining the embedding using the stable solution is not straightforward mainly due to the two-fold ambiguity of the solution. Furthermore, an integration step is also required before we can obtain depth from the disambiguated solution. We give the details of the proposed stable type-I method as follows. We propose to use $\bar{\beta}$ which is stable in

perspective and affine conditions to solve SfT. In order to obtain depth $\hat{\varphi}_z$ from $\bar{\beta}$ we need to go through the following four steps: *i)* sign disambiguation for $\bar{\beta}$, *ii)* numerical integration of $\bar{\beta}$, *iii)* arbitrary integration constant computation and *iv)* variable change, for the perspective camera. For the infinitesimal weak-perspective camera the last step is not required and we use $\bar{\phi}_z, \bar{\kappa}$ instead of $\bar{\alpha}, \bar{\beta}$.

Algorithm 1: Stable type-I method for the perspective camera.

Input: warp η, template embedding Δ, domain Ω
Output: deformed embedding $\hat{\varphi}$

- **PDE solution**
  1. Compute the Jacobians J_η and J_Δ
  2. Solve the PDE system (16) to obtain $\bar{\alpha}$ and $\pm\bar{\beta}$
- **Sign disambiguation**
  3. Compute the Jacobian $\mathsf{J}_{\bar{\alpha}}$ for the solution $\bar{\alpha}$
  4. Select the sign for $\bar{\beta}$ so that the largest component of $\bar{\beta}$ has the same sign as that component of $\mathsf{J}_{\bar{\alpha}}$
  5. Compute the two absolute angles between the vectors $\mathsf{J}_{\bar{\alpha}}$ and $\bar{\beta}$
  6. Discard $\bar{\beta}$ in points where the best angle is greater than a threshold (sum of the mean angle for all points and two times the standard deviation)
  7. Interpolate to compute $\bar{\beta}$ for the discarded values
- **Numerical integration**
  8. Integrate the disambiguated value of $\bar{\beta}$ to obtain $\hat{\alpha} + k_\alpha$
- **Integration constant**
  9. Compute the integration constant:

     $$k_\alpha = \underset{\Omega}{\text{median}}\left((\hat{\alpha} + k_\alpha) - \bar{\alpha}\right)$$

- **Change of variable**
  10. Apply the change of variable $\hat{\varphi}_z = \frac{\hat{\alpha}}{\nu}$ and compute $\hat{\varphi}$ using the perspective camera model

Algorithm 2: Stable type-I method for the infinitesimal weak-perspective camera.

Input: warp η, template embedding Δ, domain Ω
Output: deformed embedding $\hat{\varphi}$

- **PDE solution**
  1. Compute the Jacobians J_η and J_Δ
  2. Solve the PDE system (23) to obtain $\bar{\varphi}_z$ and $\pm\bar{\kappa}$
- **Sign disambiguation**
  3. Compute the Jacobian $\mathsf{J}_{\bar{\varphi}_z}$ for the solution $\bar{\varphi}_z$
  4. Select the sign for $\bar{\kappa}$ so that the largest component of $\bar{\kappa}$ has the same sign as that component of $\mathsf{J}_{\bar{\varphi}_z}$
  5. Compute the two absolute angles between the vectors $\mathsf{J}_{\bar{\varphi}_z}$ and $\pm\bar{\kappa}$
  6. Discard $\bar{\kappa}$ in points where the best angle is greater than a threshold (sum of the mean angle for all points and two times the standard deviation)
  7. Interpolate to compute $\bar{\kappa}$ for the discarded values
- **Numerical integration**
  8. Integrate the disambiguated value of $\bar{\kappa}$ to obtain $\hat{\varphi}_z + k_z$
- **Integration constant**
  9. Compute the integration constant:

     $$k_z = \underset{\Omega}{\text{median}}\left((\hat{\varphi}_z + k_z) - \bar{\varphi}_z\right)$$

  10. Find the embedding $\hat{\varphi}$ using the perspective camera model

1) Sign disambiguation: According to equation (21), non-holonomic solutions $\bar{\beta}$ and $\bar{\kappa}$ are known up to a local sign change. [Bartoli et al., 2015; Pizarro et al., 2013] mention a few ways to disambiguate the sign for different but related problems, based on external cues, such as shading, temporal smoothing, or surface smoothing. We show below that we can do without these additional cues, which may be unavailable or even unstable in practice.

If there is some perspective, even very loose, we know that a non-holonomic solution for φ_z exists. We thus propose to disambiguate the sign of $\bar{\beta}$ or $\bar{\kappa}$ by using the non-holonomic solution to depth $\bar{\varphi}_z$. In the perspective camera the process has four steps: **1**) We first differentiate $\bar{\alpha}$ to obtain $\mathsf{J}_{\bar{\alpha}}$. **2**) We select the sign of $\bar{\beta}$ so that the resulting vector is closest to $\mathsf{J}_{\bar{\alpha}}$. **3**) We discard the computed $\bar{\beta}$ at regions of the template where $\mathsf{J}_{\bar{\alpha}}$ differs substantially from $\bar{\beta}$. This can occur due to the instability of the depth solution. We use the angle between the two vectors as a metric:

$$\measuredangle(\mathbf{p}) = \left| \operatorname{acos}\left(\frac{\mathsf{J}_{\bar{\alpha}}\bar{\beta}}{\|\mathsf{J}_{\bar{\alpha}}\|\|\bar{\beta}\|} \right) \right|. \tag{36}$$

The above computed angle is simply an angle between vectors and does not have a physical significance. It is only used as a metric to choose among the solutions $\pm\bar{\beta}$. We use the sum of the mean angle for all points and twice the standard deviation as the threshold. **4**) We use smoothing to compute values for $\bar{\beta}$ for the regions where they were discarded.

2) Numerical integration: The non-holonomic solution $\bar{\beta}$ is not guaranteed to be integrable. We thus need a numerical integration method to estimate $\hat{\varphi}_z$. We propose to use a parametric function represented by a Bicubic B-Spline (BBS). With a BBS, or any other linear basis expansion model, we can integrate $\bar{\beta}$ by means of sparse linear least-squares. The solution is defined up to an additive integration constant. We define the LLS integration as:

$$\hat{\alpha} + k_\alpha = \underset{\alpha_s}{\operatorname{argmin}} \int_\Omega \|\mathsf{J}_{\alpha_s} - \bar{\beta}\|^2 \, d\mathbf{p}. \tag{37}$$

We evaluate the integral by using a summation over a dense grid of points $\mathbf{p}$ on the flat template space Ω.

3) Integration constant: After integration we obtain $\hat{\alpha} + k_\alpha$, where k_α is an arbitrary integration constant. We propose to use $\bar{\alpha}$ to estimate k_α. First we take samples from $\hat{\alpha} + k_\alpha$ and $\bar{\alpha}$. We then obtain k_α by using the median of the differences between the samples. This is expressed as:

$$k_\alpha = \underset{\Omega}{\operatorname{median}} \left((\hat{\alpha} + k_\alpha) - \bar{\alpha} \right). \tag{38}$$

4) Change of variable: We apply the change of variable given by equation (15) on the estimate of $\hat{\alpha}$ obtained above. This gives us the depth estimate at each point and thus the final surface embedding is obtained by estimating the surface point coordinates using the perspective camera model. Algorithm 1 summarizes the steps for the stable type-I method for the perspective camera.

For the infinitesimal weak-perspective camera, all the steps are very similar except that we directly obtain $\bar{\varphi}_z$ and $\bar{\kappa}$ as the non-holonomic solutions instead of $\bar{\alpha}$ and $\bar{\beta}$ respectively. This means that while the rest of the steps remain the same, the final depth and shape are obtained without the change of variable. For clarity, we describe these steps separately for the infinitesimal weak-perspective camera in algorithm 2.

V. Type-II Solutions, Stability and Type-II Stable Method

The type-I solutions discussed in §IV involve a change of variable of the depth and its derivatives where the non-holonomic solutions are measured in a modified space for the perspective camera. We present a new interpretation of SfT using the type-II solutions and type-II stable methods for two important reasons. First type-II PDE for SfT has not been studied before; it describes explicitly how tangent planes on the embedding are related by rigid transforms from the template to the deformed surface. Additionally, as we show in the experimental results, it can lead to slightly different results from the stable type-I method. This is primarily due to the use of different spaces for numerical integration, which is further explained in section VII and Appendix B. In this section we propose non-holonomic solutions of the general PDE (12) involving depth and the surface normal, which provide a more intuitive geometrical interpretation despite being equivalent to type-I solutions up to a change of variable.

A. Type-II Solutions

To use the PDE system proposed in [Collins and Bartoli, 2014a] for SfT, we need to ensure that the template shape is locally planar with normals pointing towards the positive depth axis. Such case is easily realized for a flat template. For a general 3D template, we make use of a locally isometric flattening operator, which provides a new flat parametrization space that is related to the 3D template by a locally isometric map. We then exploit the fact that isometric deformations induce a rigid transformation between the tangent plane in the template and the one in the surface embedding. Thus, from the deformation constraint (4), we have that the embedding's Jacobian is a Stiefel matrix. We use this property to find the non-holonomic solutions. For general 3D templates, the flattening is not necessarily an isometric map. Therefore we first show that one can always change the parametrization space so that locally the template parametrization (flattening) is isometric. We call the new parametrization space as the locally isometric flattening. This change involves computing the Cholesky factorization of $\mathsf{J}_{\Delta}^{\top}\mathsf{J}_{\Delta}$ at every point. We first present the locally isometric flattening and then describe the new set of non-holonomic solutions for the depth and the surface normal.

1) Locally isometric flattening: For a generic flattening with parametrization Δ, the columns of the embedding's Jacobian J_{φ} are in general not orthonormal. However we prove here that we can still get a local embedding using a different parametrization space such that the embedding's Jacobian will have orthonormal columns. We consider the required local embedding as $\phi_{\mathbf{p}'} \in C^1(\Omega', \mathbb{R}^3)$ parametrized with a new flat template space Ω' such that the new flat template is locally isometric to the 3D template $\mathcal{T}$. This leads us to the following equation:

$$\mathsf{J}_{\phi_{\mathbf{p}'}}^{\top}\mathsf{J}_{\phi_{\mathbf{p}'}} = \mathsf{J}_{\Delta'_{\mathbf{p}'}}^{\top}\mathsf{J}_{\Delta'_{\mathbf{p}'}} = \mathbf{I}_2 \tag{39}$$

where $\Delta'_{\mathbf{p}'} \in C^1(\Omega', \mathbb{R}^3)$ is a locally isometric parametrization that uses the new flat template and maps each point on Ω' isometrically to the 3D template $\mathcal{T}$. Here the domain Ω' of $\Delta'_{\mathbf{p}'}$ in general is not a connected space. We also map each point on the original flat template space Ω to the new flat template space Ω' using a local function $\rho_{\mathbf{p}} \in C^1(\Omega, \mathbb{R}^2)$. As it is a one to one mapping between subsets of $\mathbb{R}^2$, its inverse $\rho_{\mathbf{p}'}^{-1} \in C^1(\Omega', \mathbb{R}^2)$ is well defined.

We show the complete parametrization in figure 2. In the following discussion, we show that such a template can be constructed for any surface embedded in $\mathbb{R}^3$ in a disc topology. To simplify the notations, we drop the subscripts $\mathbf{p}$ and $\mathbf{p}'$ from the local functions.

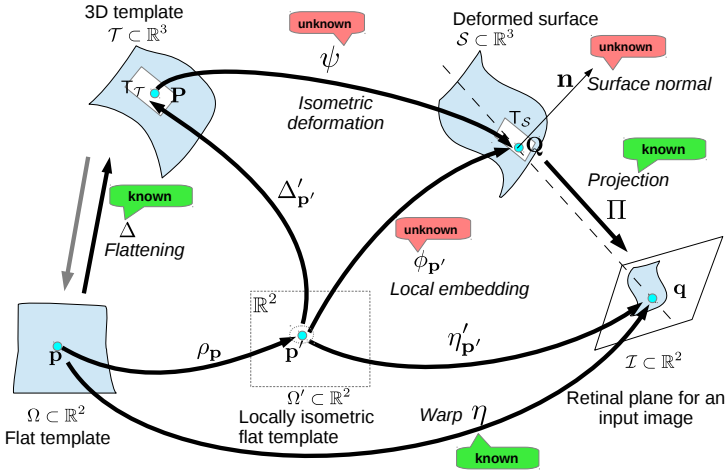


Fig. 2: Differential geometric modeling of Shape-from-Template with the locally isometric flattening. The new space depicted in the middle is locally isometric to the 3D template and the deformed surface. This property is required to construct the type-II SfT PDE.

The only new restriction that needs to be imposed for the new flattening is equation (39). Δ' can be expressed in terms of Δ and ρ^{-1} as $\Delta' = \Delta \circ \rho^{-1}$. This gives us the first-order relation:

$$\mathsf{J}_{\Delta'} = (\mathsf{J}_\Delta \circ \rho^{-1})\mathsf{J}_{\rho^{-1}}, \tag{40}$$

where $\mathsf{J}_\Delta \circ \rho^{-1}$ is the Jacobian of the known parametrization function of the 3D template evaluated in Ω. Combining equations (39) and (40) we obtain:

$$\mathsf{J}_{\rho^{-1}}^{-\top}\mathsf{J}_{\rho^{-1}}^{-1} = (\mathsf{J}_\Delta \circ \rho^{-1})^\top(\mathsf{J}_\Delta \circ \rho^{-1}). \tag{41}$$

Our goal here is to find the Jacobian of the new template-to-image warp $\mathsf{J}_{\eta'}$ and for that purpose it suffices for us to compute the Jacobian $\mathsf{J}_{\rho^{-1}}$. Any ρ^{-1} that satisfies equation (41) will lead to a $\mathsf{J}_{\Delta'}$ which will in turn satisfy equation (39). As equation (39) is the only requirement for the locally isometric template, finding $\mathsf{J}_{\rho^{-1}}$ is equivalent to obtaining the new parametrization. To obtain a value for $\mathsf{J}_{\rho^{-1}}$ that is consistent with equation (41) we perform the Cholesky decomposition of the right-hand side of equation (41) which is a symmetric positive definite matrix. This gives us:

$$\text{Chol}\left(\left(\mathsf{J}_\Delta \circ \rho^{-1}\right)^\top \left(\mathsf{J}_\Delta \circ \rho^{-1}\right)\right) = \chi\chi^\top$$

where $\chi \in C^0(\Omega', GL_2)$ is a lower triangular matrix-valued function. A value for $\mathsf{J}_{\rho^{-1}}$ that satisfies equation (41) is:

$$\mathsf{J}_{\rho^{-1}} = \chi^{-\top}. \tag{42}$$

There is in fact a class of matrices that differ by a single rotation, which satisfy equation (41). For all purposes, the value of the Jacobian $\mathsf{J}_{\rho^{-1}}$ given by equation (42) corresponds to a valid ρ^{-1}. We proceed further by considering the

new flat template-to-image warp as $\eta' : \Omega' \to \mathbb{R}^2$. η' is related to the known flat template-to-image warp η as:

$$\eta' = \eta \circ \rho^{-1}. \tag{43}$$

Thus the Jacobian of the new template-to-image warp η' is obtained as:

$$\mathsf{J}_{\eta'} = \left(\mathsf{J}_{\eta} \circ \rho^{-1}\right) \mathsf{J}_{\rho^{-1}}. \tag{44}$$

2) SfT PDE with locally isometric flattening: After changing the parametrization space to a locally isometric flattening, we are now ready to give the analytical solutions for the depth ϕ_z and the surface normal $\mathbf{n} \in C^0(\Omega', \mathbb{R}^3)$ for each point on $\mathcal{S}$. Note that the new flat template space implies no changes in the function values of the embedding or the template-to-image warp but their derivatives are however different from those of the original functions. Thus we have $\eta(\mathbf{p}) = \eta'(\mathbf{p}')$ and $\varphi(\mathbf{p}) = \phi(\mathbf{p}')$ but $\mathsf{J}_{\eta'} \neq \mathsf{J}_{\eta}$ and $\mathsf{J}_{\phi} \neq \mathsf{J}_{\varphi}$ in general. The new deformation constraint is given as:

$$\mathsf{J}_{\phi}^{\top} \mathsf{J}_{\phi} = \mathbf{I}_2. \tag{45}$$

Considering the change in the parametrization space, the new reprojection constraint in the PDE system (12) becomes:

$$\Pi \circ \phi = \eta'. \tag{46}$$

Differentiating equation (46) we obtain:

$$\left(\mathsf{J}_{\Pi} \circ \phi\right) \mathsf{J}_{\phi} = \mathsf{J}_{\eta'}. \tag{47}$$

We write the generalized type-II PDE system for SfT by combining equations (45) and (47) as:

$$\text{Find } \phi_z,\ \mathsf{J}_{\phi} \text{ s.t.} \quad \begin{cases} \left(\mathsf{J}_{\Pi} \circ \phi\right) \mathsf{J}_{\phi} = \mathsf{J}_{\eta'} \\ \mathsf{J}_{\phi}^{\top} \mathsf{J}_{\phi} = \mathbf{I}_2. \end{cases} \tag{48}$$

We select perspective or infinitesimal weak-perspective systems by substituting the value of $\mathsf{J}_{\Pi} \circ \phi$ accordingly in the PDE system (48).

3) Perspective camera: We parametrize the surface 3D points on $\mathcal{S}$ using the depth function ϕ_z as $\phi = \phi_z[\frac{\eta'^{\top}}{f}\ 1]^{\top}$. This allows us to expand the Jacobian of the projection matrix evaluated on the surface as:

$$\mathsf{J}_{\Pi} \circ \phi = \frac{f}{\phi_z} \begin{bmatrix} \mathbf{I}_2 & -\frac{\eta'}{f} \end{bmatrix}. \tag{49}$$

Combining equations (47) and (49) we obtain the following reprojection constraint:

$$\frac{f}{\phi_z} \begin{bmatrix} \mathbf{I}_2 & -\frac{\eta'}{f} \end{bmatrix} \mathsf{J}_{\phi} = \mathsf{J}_{\eta'}. \tag{50}$$

With equation (50) we can rewrite our new PDE system for SfT as:

$$\text{Find } \phi_z,\ \mathsf{J}_\phi \text{ s.t. } \quad \begin{cases} \frac{f}{\phi_z}\begin{bmatrix}\mathbf{I}_2 & -\frac{\eta'}{f}\end{bmatrix}\mathsf{J}_\phi = \mathsf{J}_{\eta'} \\ \mathsf{J}_\phi^\top \mathsf{J}_\phi = \mathbf{I}_2. \end{cases} \tag{51}$$

We denote the embedding's Jacobian J_ϕ as τ_{32} where $\tau \in C^0(\Omega', SO_3)$ is a function giving a rotation matrix. The third column of τ gives the solution for the surface normal $\bar{\mathbf{n}}$ and the first two columns form the embedding's Jacobian J_ϕ. If we obtain J_ϕ and thus the first two columns of τ, the surface normal $\bar{\mathbf{n}}$ can be retrieved from the cross-product of the two columns of τ_{32}. We give the steps for computing the non-holonomic solutions of system (51) below.

We first compute a rotation function $\theta \in C^0(\Omega', SO_3)$ such that:

$$\begin{bmatrix}\mathbf{I}_2 & -\frac{\eta'}{f}\end{bmatrix}\theta = \begin{bmatrix}\omega & \mathbf{0}\end{bmatrix} \tag{52}$$

where $\omega \in C^0(\Omega', \mathbb{R}^{2\times 2})$ is a matrix-valued function. θ and ω can be computed directly from equation (52). The actual steps are provided in Appendix A. Next we multiply the left-hand side of equation (50) by $\theta\theta^\top$ to obtain:

$$\frac{f}{\phi_z}\begin{bmatrix}\omega & \mathbf{0}\end{bmatrix}\theta^\top \tau_{32} = \mathsf{J}_{\eta'}. \tag{53}$$

We simplify equation (53) by introducing another rotation function $\xi \in C^0(\Omega', SO_3)$ such that $\xi_{32} = \theta^\top \tau_{32}$. This gives us the following new equation with the unknown depth ϕ_z and the unknown sub-Stiefel matrix ξ_{22}:

$$\frac{f}{\phi_z}\xi_{22} = \omega^{-1}\mathsf{J}_{\eta'}. \tag{54}$$

To solve equation (54) we use the fact that $\xi_{22} \in SS_{2\times 2}$ is a sub-Stiefel matrix and thus its largest singular value is 1. Equating the largest singular values for both sides of equation (54), we obtain:

$$\phi_z = \frac{1}{f}\sigma_1^{-1}\left(\omega^{-1}\mathsf{J}_{\eta'}\right) \tag{55}$$

where $\sigma_1(\mathbf{M})$ is an operator giving the largest singular value of the matrix $\mathbf{M}$. Similarly ξ_{22} can be recovered from the relation $\xi_{22} = \frac{1}{f}\phi_z\omega^{-1}\mathsf{J}_{\eta'}$.

We parametrize the Stiefel matrix ξ_{32} by using the computed sub-Stiefel part ξ_{22} and an unknown vector $\mathbf{r}$ so that $\xi_{32}^\top = [\xi_{22}^\top\ \mathbf{r}]$. Using the orthogonality of the Stiefel matrix we have:

$$\mathbf{I}_2 - \xi_{22}^\top\xi_{22} = \mathbf{r}\mathbf{r}^\top. \tag{56}$$

Since $\mathbf{r}$ is a vector, both sides of equation (56) have rank 1. Consequently, $\mathbf{r}$ can be obtained by taking the SVD of the left-hand side of equation (56) and choosing the right singular vector corresponding to its non-zero singular value as below:

$$\mathbf{r} = \pm\mathbf{v}_1(\mathbf{I}_2 - \xi_{22}^\top\xi_{22}). \tag{57}$$

However, this results in two solutions for ξ_{32}; we write them as: ξ_{32}^a and ξ_{32}^b. The two solutions for the embedding's

Jacobian are now simply given by:

$$\mathsf{J}_\phi = \tau_{32} = \begin{cases} \theta\xi^a_{32} \\ \theta\xi^b_{32}. \end{cases} \tag{58}$$

Finally we can also obtain the two solutions for the surface normal $\bar{\mathbf{n}}_1$ and $\bar{\mathbf{n}}_2$ from the cross product of the columns of the two solutions for τ_{32}.

4) Infinitesimal weak-perspective camera: The infinitesimal weak-perspective camera model can be used in place of the perspective model by parametrizing the 3D image points with the infinitesimal weak-perspective depth function ζ and the template-to-image warp in homogeneous coordinates $\tilde{\eta}' = [\eta'^\top \ 1]^\top$:

$$\phi = \frac{\zeta}{f}\tilde{\eta}'. \tag{59}$$

Here ζ varies over the template while its gradient $\mathsf{J}_\zeta = \mathbf{0}$. This gives us the Jacobian of the projection matrix evaluated on the surface as:

$$\mathsf{J}_\Pi \circ \phi = \frac{f}{\zeta}\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix}. \tag{60}$$

Allowing $\zeta = \phi_z$, *i.e.* to be different for each point while keeping the Jacobian of the projection matrix as in equation (60) gives us the infinitesimal weak-perspective system. Thus combining equations (47) and (60) we obtain the following reprojection constraint for the infinitesimal weak-perspective model:

$$\frac{f}{\phi_z}\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix}\mathsf{J}_\phi = \mathsf{J}_{\eta'}. \tag{61}$$

As in the perspective solutions we form the PDE system by considering J_ϕ as a Stiefel matrix. Thus the problem becomes:

$$\text{Find } \phi_z,\ \mathsf{J}_\phi \text{ s.t. } \quad \frac{f}{\phi_z}\begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix}\tau_{32} = \mathsf{J}_{\eta'}, \tag{62}$$

where $\tau \in C^1(\Omega', SO_3)$ is a rotation matrix such that $\mathsf{J}_\phi = \tau_{32}$. Simplifying problem (62) leads to the following:

$$\frac{f}{\phi_z}\tau_{22} = \mathsf{J}_{\eta'}. \tag{63}$$

Equation (63) is identical to the perspective system described by equation (54) except that we do not have a rotation θ involved in the infinitesimal weak-perspective system and thus the equation is much simpler. As we are allowing ϕ_z to change at each point and solving for the pose and scale simultaneously the solutions are infinitesimal weak-perspective solutions. We obtain ϕ_z as:

$$\phi_z = \frac{1}{f}\sigma_1^{-1}\left(\mathsf{J}_{\eta'}\right). \tag{64}$$

Similarly τ_{22} can be obtained as:

$$\tau_{22} = \frac{1}{f}\phi_z\mathsf{J}_{\eta'}. \tag{65}$$

We recover the two solutions for τ_{32} in the same way as we did for ξ_{32} for the perspective camera and thus obtain the non-holonomic solutions for the surface normal using the cross product.

5) Obtaining the embedding: Solving the generalized type-II SfT PDE system (48) gives us the set of non-holonomic solutions: $\bar{\phi}_z$ and $\bar{\mathbf{n}}_{1,2}$. One obvious way to obtain the embedding ϕ is to use the direct depth solution $\bar{\phi}_z$ and the normalized points obtained from η as: $\bar{\phi} = \bar{\phi}_z[\frac{\eta'^\top}{f}\ 1]^\top$. This solution is identical to the direct-depth method proposed in [Bartoli et al., 2015] as both of them are the non-holonomic solutions for the depth. However in doing so, we essentially discard the solution for the Jacobian or the normal and make use of only one non-holonomic solution *i.e.* the depth ϕ_z. In §V-B we describe how the solution for the Jacobian $\bar{\tau}_{32}$ and thus the normal $\bar{\mathbf{n}}$ is better constrained than the depth solution $\bar{\phi}_z$. We provide the actual algorithm for obtaining a stable 3D reconstruction with the stable type-II method in §V-C.

B. Stability

We show here that the amount of perspective affects how depth is constrained in the type-II solutions in the same way as in the type-I solutions. We prove two important results for the type-II solutions.

Proposition 3. *The non-holonomic solution for depth $\bar{\phi}_z$ is weakly constrained when the projection geometry tends to affine.*

Proposition 4. *The non-holonomic solution for the embedding's Jacobian and thus the surface normal $\bar{\mathbf{n}}$ is well-constrained in all projection geometries.*

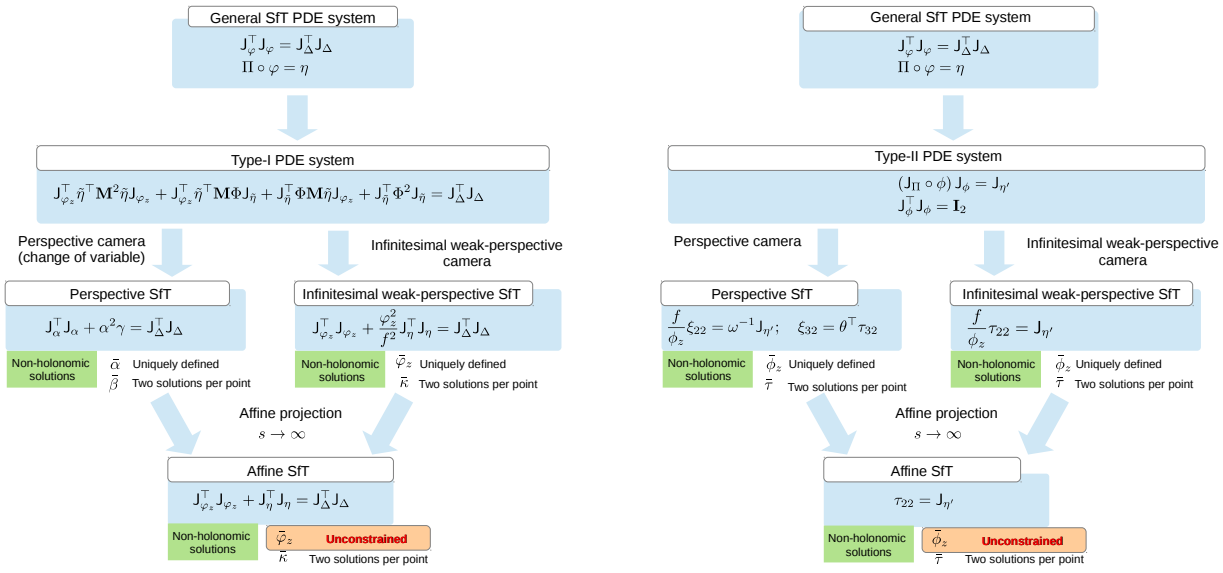


Fig. 3: SfT type-I solutions (left) and type-II solutions (right) for different projection models and amount of perspective.

Figure 3 illustrates the effect of projection geometries on the *type-I* and *type-II solutions* of SfT. As in §IV-B we make use of the projection function Π_s that depends on a parameter s and the initial focal length f to continuously select the amount of perspective. The only difference being that we now use the symbol ϕ instead of φ for the embedding.

Proof of propositions 3 and 4 for the perspective camera. We take the first-order derivatives of equation (26) on the surface, which gives us:

$$\mathsf{J}_{\Pi_s}(\mathbf{Q}) = f\frac{s+1}{Q_z+sf}\begin{bmatrix}\mathbf{I}_2 & \frac{-\eta'}{f(s+1)}\end{bmatrix}. \tag{66}$$

Combining equation (47) with equation (66), we obtain:

$$f\frac{s+1}{Q_z+sf}\begin{bmatrix}\mathbf{I}_2 & \frac{-\eta'}{f(s+1)}\end{bmatrix}\tau_{32} = \mathsf{J}_{\eta'}. \tag{67}$$

We first prove proposition 3. Evaluating the limit $s \to \infty$ for equation (67) gives us:

$$\begin{bmatrix}\mathbf{I}_2 & 0\end{bmatrix}\tau_{32} = \mathsf{J}_{\eta'} \quad \Leftrightarrow \quad \tau_{22} = \mathsf{J}_{\eta'}. \tag{68}$$

The depth Q_z is now no longer constrained in equation (68) at the limit. This proves our first result that depth is not constrained for affine projection in SfT.

The proof of proposition 4 follows directly from the fact that τ_{22} is well-constrained in equation (68) despite the projection geometry. Following the steps in [Collins and Bartoli, 2014a] or in §V-A, one can easily derive the two solutions for J_ϕ and thus the two solutions for the normal from equation (68). □

Proof of propositions 3 and 4 for the infinitesimal weak-perspective camera. Here we start with the infinitesimal weak-perspective approximation for the projection function given by equation (28). Taking the limit $s \to \infty$ in equation (28) leads to an expression identical to an expression identical to equation (68). The proof for propositions 3 and 4 then follows with the same arguments as in the perspective camera. □

C. Stable Type-II Methods

We presented the non-holonomic type-II solutions for the general SfT PDE system (12) and proved that the depth solution was unstable. We give the method for obtaining the final embedding using the stable solution as follows. The stable type-II methods use the solution to the surface normal directly to obtain the final embedding. We first obtain the surface normal by taking the cross product of the Jacobian columns and then to use Shape-from-Normals to obtain depth and the final embedding. The two steps are almost identical but having an analytical surface normal gives a better geometrical appeal to the problem. Furthermore, one can imagine several scenarios where the required end result from SfT would be a surface normal rather than the embedding itself. However, the non-holonomic solution to the embedding's Jacobian or the normal have a two-fold ambiguity. Therefore we solve the following problems in our type-II methods to obtain the 3D embeddding: *i)* normal disambiguation, *ii)* Shape-from-Normals and *iii)* scale computation. All of these steps are identical for the perspective and infinitesimal weak-perspective cameras. Figure 4 illustrates the specifics and similarities of stable type-I and stable type-II methods together.

1) Normal disambiguation: In type-II solutions we obtain two different non-holonomic solutions for the embedding's Jacobian τ_{32} and thus two different values for the normal vector: $\bar{\mathbf{n}}_1$ and $\bar{\mathbf{n}}_2$. To disambiguate the surface normal, we go through the following list of steps. **1**) We first compute a surface embedding $\bar{\phi}$ from the direct depth solution and a

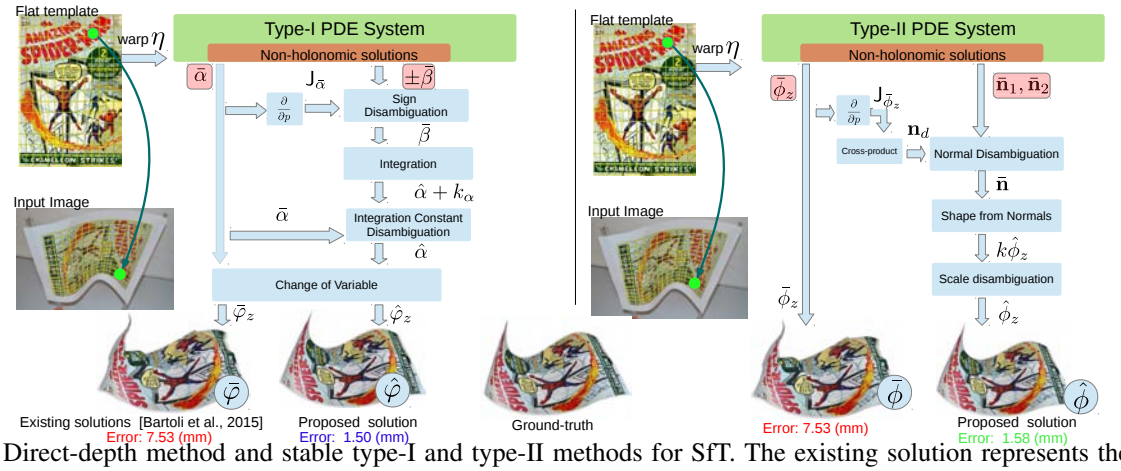


Fig. 4: Direct-depth method and stable type-I and type-II methods for SfT. The existing solution represents the results from the direct-depth method and the proposed solutions represent the results from the stable methods.

normal field $\mathbf{n}_d$ using the surface $\bar{\phi}$. **2**) For each point on the surface, we compute the following dot product with the normal vector $\mathbf{n}_d$: $|\mathbf{n}_d^\top \bar{\mathbf{n}}_1|$ and $|\mathbf{n}_d^\top \bar{\mathbf{n}}_2|$. **3**) We select the highest dot product among the two and choose the corresponding normal vector $\bar{\mathbf{n}}_1$ or $\bar{\mathbf{n}}_2$. This process can be expressed as: $\bar{\mathbf{n}} = \bar{\mathbf{n}}_k$, such that, $k = \text{argmax}_{k\in\{1,2\}} |\mathbf{n}_d^\top \bar{\mathbf{n}}_k|$. Unlike in the type-I methods we do not remove regions or points here, as it gives little or no improvement.

2) Shape-from-Normals: Once we disambiguate the normals at each point, the next step is to obtain an embedding by integrating the normals. This can be done very efficiently in a least-squares manner using spline functions such as the B-splines. With the integration we obtain depth at each evaluated point. The integration of normals is defined as:

$$\hat{\phi}_k = k_z \hat{\phi} = \underset{\phi_s}{\text{argmin}} \int_{\Omega'} \left(\left(\bar{\mathbf{n}}^\top [\mathsf{J}_{\phi_s}]_1\right)^2 + \left(\bar{\mathbf{n}}^\top [\mathsf{J}_{\phi_s}]_2\right)^2 \right) d\mathbf{p}' \tag{69}$$

where k_z represents the unknown scale of reconstruction due to the integration of unit normals and $\hat{\phi}_k \in C^1(\Omega', \mathbb{R}^3)$ is the shape obtained after integration and $\mathbf{p}'$ represents a point on Ω'. We represent the jth column of the Jacobian matrix J_{ϕ_s} as $[\mathsf{J}_{\phi_s}]_j$.

3) Scale computation: The surface embedding obtained by integrating the unit surface normals are not in the correct scale. To fix the scale, we first compute an approximate embedding $\bar{\phi}$ from the direct-depth solution. We parametrize the obtained shape as $\hat{\phi}_k = [Q_x \; Q_y \; Q_z]^\top$. We then use $\bar{\phi}$ and the shape $\hat{\phi}_k$ obtained from Shape-from-Normals to fix the scale as follows:

$$k_z = \left(\frac{\int_{\Omega'} \left(Q_x \bar{\phi}_x + Q_y \bar{\phi}_y + Q_z \bar{\phi}_z \right) d\mathbf{p}'}{\int_{\Omega'} \left(Q_x^2 + Q_y^2 + Q_z^2 \right) d\mathbf{p}'} \right)^{-1}. \tag{70}$$

Thus after the scale correction we obtain a stable reconstruction from the stable type-II method. These steps are summarized in algorithm 3.

VI. Experimental Results

A. Compared Methods and Error Measurements

We performed the experiments using MATLAB. The results of the experiments were used to obtain plots for the depth error (3D error) and the normal error in different conditions. We computed the depth error by taking the root mean

Algorithm 3: Stable type-II methods for the perspective and infinitesimal weak-perspective camera.

Input: warp η, template embedding Δ, domain Ω
Output: deformed embedding $\hat{\phi}$

- **PDE Solution**
  1. Compute the Jacobians J_η and J_Δ
  2. Change the parametrization space and compute $\mathsf{J}_{\eta'}$
  3. Solve the PDE system (51) or (62) to obtain $\bar{\phi}_z$ and $\bar{\mathbf{n}}_1, \bar{\mathbf{n}}_2$
- **Normal disambiguation**
  4. Compute a surface embedding $\bar{\phi}$ from the direct depth solution $\bar{\phi}_z$
  5. Find a normal field $\mathbf{n}_d$ on the surface using the embedding $\bar{\phi}$
  6. Select the analytical normal that is closest in angle to $\mathbf{n}_d$ corresponding to the highest dot-product, *i.e.* $\bar{\mathbf{n}} = \bar{\mathbf{n}}_k$, such that $k = \operatorname{argmax}_{k\in\{1,2\}} |\mathbf{n}_d^\top \bar{\mathbf{n}}_k|$.
- **Shape-from-Normals**
  7. Integrate the disambiguated surface normal field $\bar{\mathbf{n}}$ by solving the minimization problem (69)
- **Scale computation**
  8. Correct the scale of the surface embedding to obtain $\hat{\phi}$

square error of the reconstructed 3D point coordinates. We measured the normal error by taking the root mean square deviation in angle of the reconstructed surface normals from the ground-truth surface normals. Our set of proposed methods consists of the stable type-I method (**typeI**-P, **typeI**-WP) and the stable type-II method (**typeII**-P, **typeII**-WP). The suffixes 'P' and 'WP' stand for the perspective and infinitesimal weak-perspective camera models respectively. We ran the methods based on the infinitesimal weak-perspective camera model only for the scenes with changing focal length as they are expected to work only for large focal lengths. We use **direct**-P for the analytical direct-depth method [Bartoli et al., 2015]. We also compared against the zeroth-order methods based on inextensibility, denoted here as **inext-mdh**-P [Salzmann and Fua, 2011] and **inext-lap**-P [Ngo et al., 2015]. Finally we tested the statistically optimal cost optimization [Brunet et al., 2014] with the direct depth solution as input. We denote the refined solution as **refined**-P. We tuned each method with the best set of parameters for each dataset. We describe the complete algorithm for SfT, experimental setup and the results on each dataset separately below.

B. Complete algorithm

The SfT framework used for the analytical methods requires a 3D template and template-to-image registration warps. Here we briefly list the steps for inferring the deformed shape starting from the 3D template and an input image for real datasets where outliers appear naturally: *1)* We compute the Δ parametrization from the template flattening to the 3D template using Bicubic B-Splines (BBS) with a smoothness prior [Dierckx, 1993]. This is an LLS problem. When the 3D template is not flat, we make use of a template image as the flattening. *2)* We obtain point correspondences between the template image and the input image using SIFT [Lowe, 2004] or KAZE [Alcantarilla et al., 2012] or a combination of them. An alternative approach is to match using a denser graph matching method such as [Collins et al., 2014]. For most real datasets, these correspondences will in general contain some outliers. In either case, outliers are removed using [Pizarro and Bartoli, 2012]. *3)* The correspondences thus established will have none or very few outliers even in difficult conditions. In more difficult situations such as the example shown in Appendix C, we opt

for a convex L1-minimization in place of the LLS problem in [Dierckx, 1993] to reduce the effect of outliers while estimating the template-to-image warp. If the registrations are still not good enough, a robust M-estimator may be used but we found an L1 estimation to be sufficient for the examples. Optionally methods such as the pixel intensity based registration refinement also given in [Pizarro and Bartoli, 2012] or the affine transform based outlier rejection [Puerto and Mariottini, 2013] can be used to improve the registration if necessary. *4)* We use the registration obtained in the above step to generate 2D correspondences and the registration derivatives. *5)* Finally we obtain the reconstructed points from SfT.

C. Synthetic Data

We simulated 10 different surfaces generated by isometric deformation of a flat template surface [Perriollat and Bartoli, 2013]. The template size used was 640 px $\times$ 480 px. The images for each deformation were taken using a virtual pin-hole camera of varying focal length. We fixed the focal length using a single parameter as: $f = (s+1)500$ px. While changing the focal length, we also translated the object so that the size of its projection remained fixed in the image. In the experiments we varied s from 0 to 8. The number of correspondences N used to estimate the warp η was varied from 50 to 300. We added Gaussian noise in the images with a standard deviation σ varying between 0 and 2.4 px. In each experiment we changed only one parameter, fixing the others. The default values of the parameters were $s = 1$, $N = 100$ and $\sigma = 1.0$ px. The resulting plots from the experiments are shown in figure 5.

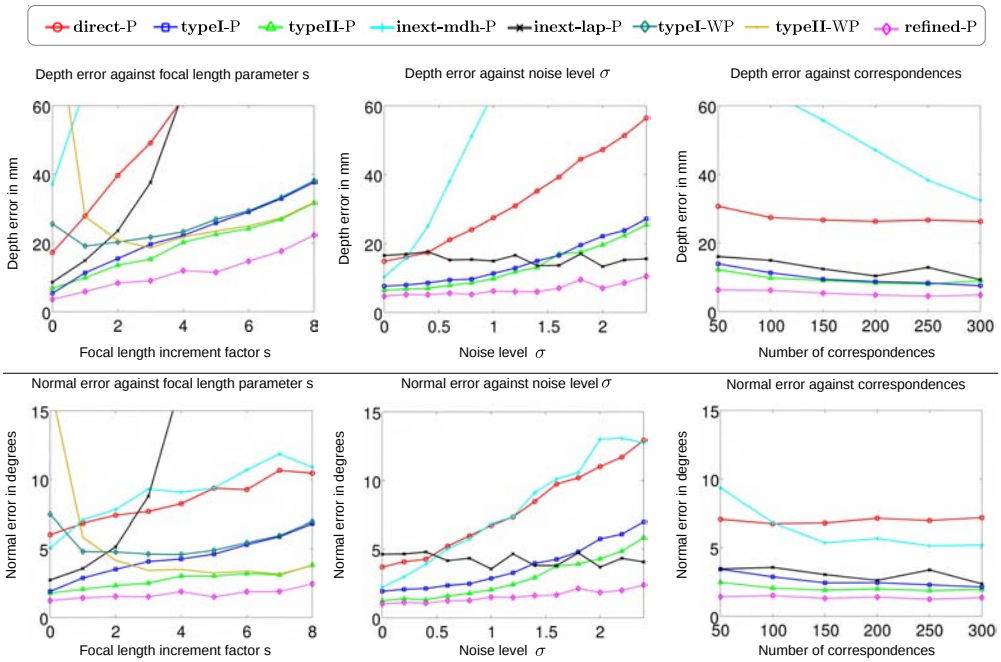


Fig. 5: Plots for the synthetic dataset. We show the depth errors in the first row and the normal errors in the second row.

The plots show that the stable methods **typeI**-P and **typeII**-P have the best performance among all methods, after **refined**-P. The stable method **typeII**-P has a slight edge over **typeI**-P in most conditions. This is because compared to **typeI**-P, **typeII**-P uses integration in a more natural space. Even so, both of these methods show better performance

than the original analytical solution **direct**-P and also compared to the zeroth-order methods **inext-mdh**-P and **inext-lap**-P. Both the direct-depth method and zeroth-order methods perform poorly against the increasing focal length. On the other hand our proposed methods **typeI**-P and **typeII**-P show remarkable stability and have performance similar to **refined**-P. Another interesting observation is the convergence of the infinitesimal weak-perspective camera model to the perspective accuracy with the increase in the focal length. These observations validate the theoretical results we obtained in §IV-B and §V-B. The zeroth-order method **inext-lap**-P performs well against the increasing number of correspondences and the image noise. The analytical solution **direct**-P shows poor performance against increasing noise while, on the other hand, both **typeI**-P and **typeII**-P are robust against noise in correspondences. Similar observations also hold for the normal error.

D. Real Data

We tested all methods with four different real datasets. We used one more dataset to make an experiment where we show how a robust registration can help the reconstruction in appendix C. The first two were built with developable surfaces with a flat 3D template. We constructed the others out of non-developable surfaces. We describe each of these datasets and observations independently below. Table I summarizes the results on all real datasets.

The KINECT Paper dataset: The KINECT Paper dataset [Varol et al., 2012] consists of 191 frames taken with about the same angle and focal length of a sheet of paper being deformed. The number of matched features in each frame of the sequence is around $N = 1300$ but varies from frame to frame. The image size is 640 px $\times$ 480 px with focal length of approximately 526 px. The performance of different methods for each frame is plotted in figure 6.

As the dataset is highly perspective and has a large number of feature correspondences distributed more or less uniformly over the scenes, all of the methods perform well. The mean depth errors for our stable methods, however, are significantly lower than those for the others, including **refined**-P, as shown in table I.

The Zooming dataset: In order to test the performance of methods with varying focal length, we used a real dataset with a single deformation [Bartoli et al., 2013] and known template. The dataset shows a folded sheet of paper with different focal lengths and views. The focal length varies from 2696 px to 7875 px with an image size of 1728 px $\times$ 1552 px. Each zoom level has 7 to 10 images with different viewing angles. We computed the ground-truth from each view in camera coordinates using stereo triangulation and feature correspondences from SIFT. Figure 6 shows the computed plots for the shape and the depth errors for different methods. We also show the qualitative results with error coded texture maps for three images of different focal lengths in figure 7.

The error plots show that both **typeI**-P and **typeII**-P perform the best in the dataset while **direct**-P, **inext-mdh**-P and **inext-lap**-P do not perform very well. In particular, there could be two important reasons behind the low accuracy of **inext-lap**-P for this dataset. First the images do not have very short focal lengths and since it requires some perspective, the results can be expected to be less accurate in such conditions. Second, the optimal parameters for the method change from image to image. We used a single set of parameters for all images in the dataset. We impose smoothness for

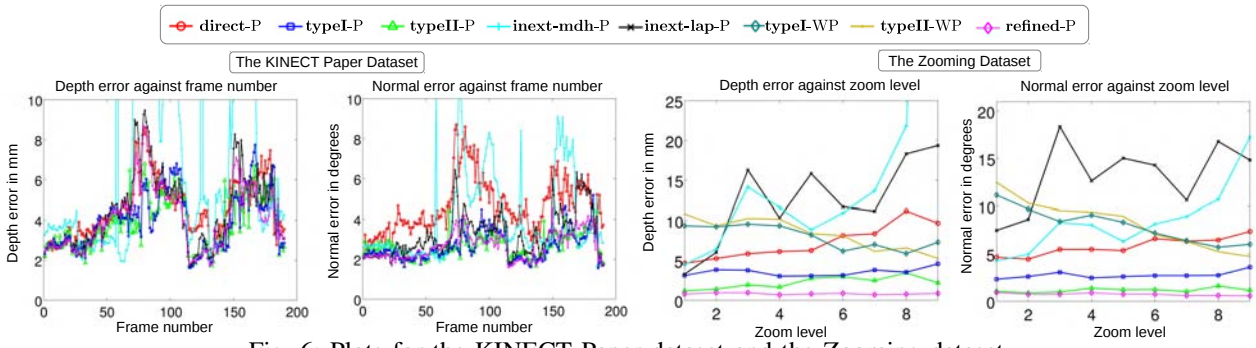
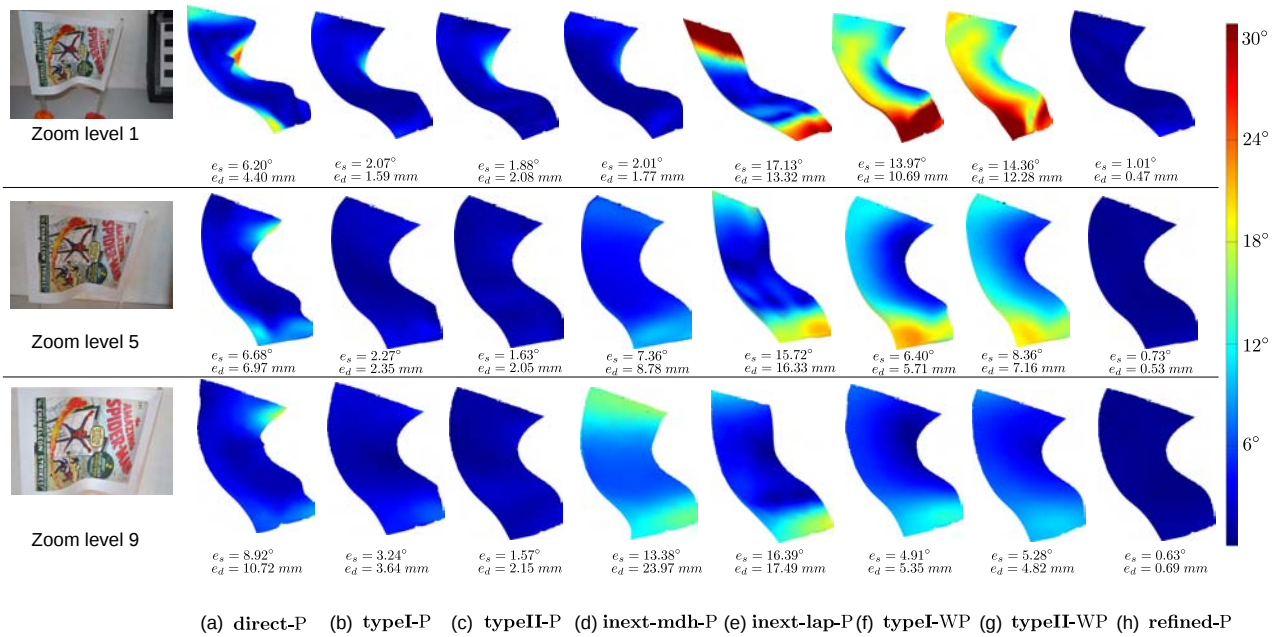


Fig. 6: Plots for the KINECT Paper dataset and the Zooming dataset.



(a) direct-P (b) typeI-P (c) typeII-P (d) inext-mdh-P (e) inext-lap-P (f) typeI-WP (g) typeII-WP (h) refined-P

Fig. 7: Rendered 3D results with error coded texture maps for the Zooming dataset. We use the normal error to generate the texture maps.

all reconstructions using a BBS warp. The results further confirm that **direct**-P is more susceptible to noise in image correspondences.

The Cushion dataset: The datasets tested so far are all developable surfaces. Thus a flat template could be used so that Δ was always an identity. A slightly different situation occurs when the template is not flat and its given flattening is non-isometric. In this dataset we allowed a cushion to deform into several shapes. The deformations were largely isometric since there was very little stretching or expansion of the textured cloth. We made 5 different deformations of the cushion and we used one of the deformation as the template. Making an isometric flattening was not possible in this case and therefore we used the template image as the flattening. This also entailed the use of a locally isometric flattening for **typeII**-P and a nontrivial Δ. The size of the images used is 3456 px $\times$ 2304 px. The focal length of the camera is about 2700 px, thus the images contain a moderately large amount of perspective.

The feature correspondences here were computed by combining SIFT and KAZE [Alcantarilla et al., 2012] features. Due to the lack of point correspondences in some regions, the computed warps have less accuracy than in all other

TABLE I: Mean depth errors in real datasets.

	Depth error measurements for different methods in mm					
Datasets	**direct**-P	**typeI**-P	**typeII**-P	**inext-mdh**-P	**inext-lap**-P	**refined**-P
KINECT Paper	4.57	3.98	**3.82**	7.78	4.43	4.02
Zooming	7.28	3.54	2.22	20.16	12.47	**0.82**
Cushion	14.37	5.02	**3.48**	7.71	7.39	5.99
Can	3.03	1.38	**1.07**	4.07	1.91	1.31

TABLE II: Compared methods and their characteristics.

Methods	Constraints	Primary computation	Time (sec)	Stability in affine condition
direct-P	Differential first-order	Small systems	0.13	Not stable
typeI-P	Differential first-order	LLS Integration	0.53	Stable
typeII-P	Differential first-order	LLS Integration	1.74	Stable
inext-mdh-P	Zeroth-order MDH	Convex optimization SOCP	2.96	Not stable
inext-lap-P	Zeroth-order inextensibility	LLS, non-convex optimization	7.18	Not stable
refined-P	Differential first-order	Non-convex optimization	26.37	Stable

datasets. We present the results for three of the deformations in figure 8 using error coded texture maps along with the shape and depth error for each reconstruction. We observe again that the stable methods **typeI**-P and **typeII**-P have the best performance. The zeroth-order methods **inext-mdh**-P and **inext-lap**-P also show good performance as the images have high perspective. However we failed to obtain good results with **direct**-P. This further confirms the greater sensitivity of **direct**-P to noise in the feature correspondences.

The Can dataset: We prepared a dataset by deforming a can made of a cardboard material. The dataset consists of 3 different deformations of different degrees and a template surface made with the original surface. As it was not possible to flatten the surface physically, we again used the template image as the flattening. The size of the images used is 4800 px $\times$ 3200 px with a focal length of 11000 px. We computed the flat template-to-image warps again by combining SIFT and KAZE feature correspondences. We present the qualitative results with error coded texture maps and error measurements in figure 8. The results show that our proposed methods are again the best performing. Similarly **direct**-P shows a medium accuracy while **inext-mdh**-P performs poorly in 2 out of 3 scenes.

VII. Discussions

Isometric SfT methods are close to achieving reconstruction accuracies that give them applicability in real scenarios. However, as we showed here, several aspects of current state-of-the-art methods, specifically their poor performance in low perspective and their sensitivity to noise in image correspondences pose major problems to achieving useful SfT results. We proposed methods that push the boundary of the state of the art further in terms of reconstruction accuracy while extending applicability in different projection geometries. We found that obtaining depth directly to get the 3D shape is not the best approach despite its appeal. In contrast, we proposed the use of stable solutions (based on depth-gradient or surface normal) that proved to be accurate in perspective as well as affine conditions. We provided theoretical proofs to their stability. Though being very similar, they differ in two important aspects: disambiguation and integration. In particular we use numerical integration with bending energy of the BBS, which imposes smoothness in the given space. Compared to the stable type-I method, the stable type-II method uses the space of surface normals

for integration that gives a slightly better performance seen in the experiments. Both of these two new methods we presented used analytical solutions and obtained near or sometimes better than the statistically optimal results. In brief, we found the first-order solution of the SfT PDE to be more stable than the zeroth-order solution.

For all the experiments, we computed the template-to-image warp required for the analytical methods globally. In future works, the warp could be computed locally. By nature, a local warp would possibly capture the local changes better than a global one, thus giving better reconstruction in cases of large local deformations. Table II shows a summary of the main characteristics for different methods. It is true that the zeroth-order methods are affected by inaccuracies of the computed warp depending on the presence of outliers but such conditions are rare in our experiments. The time noted is the average time taken to reconstruct a single scene for the real datasets using a standard desktop PC. The parameters used in our methods for the global warps are very easy to tune and the methods themselves give local solutions meaning that they can be parallelized if needed for higher speed.

Acknowledgements: We would like to thank the authors of [Salzmann and Fua, 2011] and [Ngo et al., 2015] for sharing the implementation of their methods and the authors of [Varol et al., 2012] for their dataset. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE.

REFERENCES

P. F. Alcantarilla, A. Bartoli, and A. J. Davison. KAZE features. In *ECCV*, 2012.

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

A. Bartoli, D. Pizarro, and T. Collins. A robust analytical solution to isometric shape-from-template with focal length calibration. In *ICCV*, 2013.

A. Bartoli, Y. Gérard, F. Chadebecq, T. Collins, and D. Pizarro. Shape-from-template. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37 (10):2099–2118, 2015.

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

F. Brunet, A. Bartoli, and R. Hartley. Monocular template-based 3D surface reconstruction: Convex inextensible and nonconvex isometric methods. *Computer Vision and Image Understanding*, 125:138–154, 2014.

E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

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

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

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

T. Collins and A. Bartoli. Infinitesimal plane-based pose estimation. *International Journal of Computer Vision*, 109(3):252–286, 2014a.

T. Collins and A. Bartoli. Using isometry to classify correct/incorrect 3d-2d correspondences. In *ECCV*, 2014b.

T. Collins and A. Bartoli. Realtime shape-from-template: System and applications. In *ISMAR*, 2015.

T. Collins, P. Mesejo, and A. Bartoli. An analysis of errors in graph-based keypoint matching and proposed solutions. In *ECCV*, 2014.

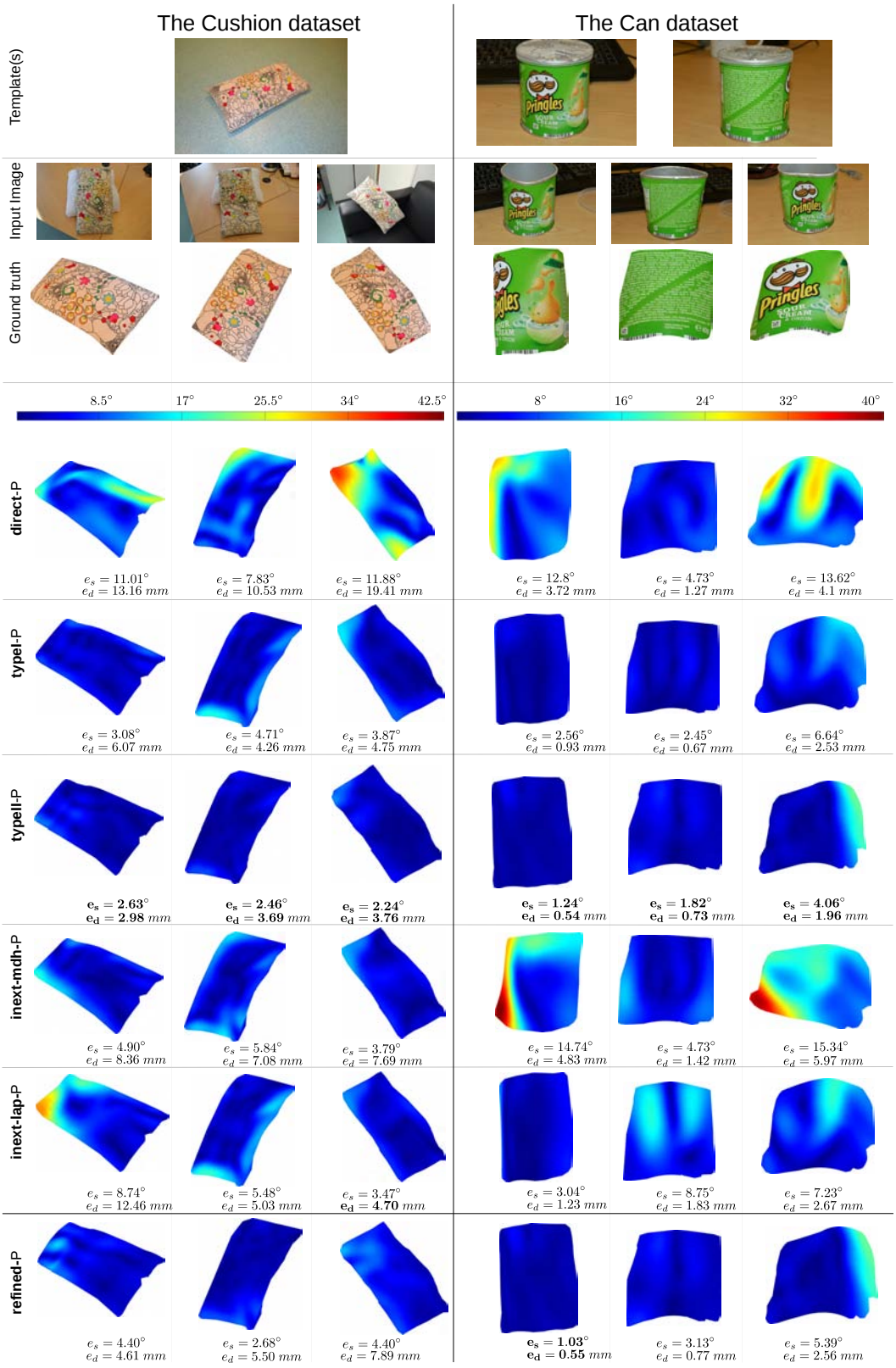


Fig. 8: Rendered 3D results with error coded texture maps for the Cushion and the Can datasets.

A. Del Bue. A factorization approach to structure from motion with shape priors. In *CVPR*, 2008.

P. Dierckx. *Curve and Surface Fitting with Splines*. Oxford University Press, Inc., 1993. ISBN 0-19-853441-8.

R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. ISBN 0521623049.

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

L. Maier-Hein, A. Groch, A. Bartoli, S. Bodenstedt, G. Boissonnat, P. Chang, N. Clancy, D. S. Elson, S. Haase, E. Heim, J. Hornegger, P. Jannin, H. Kenngott, T. Kilgus, B. P. Müller-Stich, D. Oladokun, S. Röhl, T. R. dos Santos, H. Schlemmer, A. Seitel, S. Speidel, M. Wagner, and D. Stoyanov. Comparative validation of single-shot optical techniques for laparoscopic 3-D surface reconstruction. *IEEE Transactions on Medical Imaging*, 33(10):1913–1930, 2014.

T. D. Ngo, J. O. stlund, and P. Fua. Template-based monocular 3D shape recovery using laplacian meshes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

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.

M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *International journal of computer vision*, 95(2):124–137, 2011.

J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision*, 76(2):109–122, 2008.

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

D. Pizarro, A. Bartoli, and T. Collins. Isowarp and conwarp: Warps that exactly comply with weak-perspective projection of deforming objects. In *BMVC*, 2013.

G. A. Puerto and G. L. Mariottini. A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images. *IEEE Trans. Med. Imaging*, 32(7):1201–1214, 2013.

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.

O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, 2004.

R. W. Sumner and J. Popović. Deformation transfer for triangle meshes. *ACM Transactions on Graphics*, 23(3):399–405, 2004.

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

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.

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

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

BIOGRAPHIES

Ajad Chhatkuli received his Msc degree in Computer Vision from the University of Burgundy in 2013. He is currently a PhD candidate in Computer Vision at Université d'Auvergne under Prof. Adrien Bartoli and Dr. Daniel Pizarro. His research interests include non-rigid 3D reconstruction and non-rigid registration.

Daniel Pizarro Pérez received the PhD degree in Electrical Engineering in 2008 from the University of Alcala. In 2005-2012 he was an Assistant Professor and member of the GEINTRA group at the University of Alcala. Since 2013 he is an Associate Professor at Université d'Auvergne and member of ALCoV. His research interests are in optimization and Computer Vision, including image registration and deformable reconstruction, and their application to Minimally Invasive Surgery.

Adrien Bartoli has held the position of Professor of Computer Science at Université d'Auvergne since fall 2009. He leads the ALCoV (Advanced Laparoscopy and Computer Vision) research group, member of CNRS and Université d'Auvergne, at ISIT. His main research interests include image registration and Shape-from-X for rigid and non-rigid environments, with applications to computer-aided endoscopy.

Toby Collins received the MSc degree in Artificial Intelligence at the University of Edinburgh (first in class) in 2005. In 2006 he began his PhD in Computer Vision at the University of Edinburgh. Since 2009 he has been a full-time research fellow in ALCoV. His research interests include nonrigid shape analysis, registration and reconstruction, AR for deformable surfaces and computer assisted intervention.

APPENDIX A: COMPUTATION OF θ AND ω

We need to find $\theta \in C^0(\Omega', SO_3)$ such that it satisfies equation (52):

$$\begin{bmatrix} \mathbf{I}_2 & -\frac{\eta'}{f} \end{bmatrix} \theta = \begin{bmatrix} \omega & \mathbf{0} \end{bmatrix}$$

for some $\omega \in C^0(\Omega', \mathbb{R}^{2\times 2})$. Equation (52) implies that the third column θ_3 of θ is orthonormal to both rows of $[\mathbf{I}_2 \ -\frac{\eta'}{f}]$. Note that we represent the ith column of the matrix function θ as θ_i. This gives us a closed form solution for θ_3 as:

$$\theta_3 = \frac{\tilde{\eta}'_f}{\|\tilde{\eta}'_f\|_2}, \tag{i}$$

where $\tilde{\eta}'_f \in C^1(\Omega', \mathbb{R}^3)$ is the local flat template-to-image warp function giving *normalized* homogeneous coordinates. Its components can be written as $\tilde{\eta}'_f = [\frac{\eta'_x}{f} \ \frac{\eta'_y}{f} \ 1]^\top$. The columns of θ will have the following orthonormality relations:

$$\begin{aligned} [\theta_3]_\times \theta_1 &= \theta_2 \\ [\theta_3]_\times^\top \theta_2 &= \theta_1. \end{aligned} \tag{ii}$$

Any combination of vectors θ_1 and θ_2 that satisfy equation (ii) will form the required rotation θ. First we can choose θ_1 orthogonal to θ_3 as:

$$\theta_1 = \frac{1}{\sqrt{\frac{\eta'^2_x}{f^2} + 1}} \begin{bmatrix} -1 \\ 0 \\ \frac{\eta'_x}{f} \end{bmatrix}. \tag{iii}$$

Using the values of θ_3 and θ_1, we obtain θ_2 using $\theta_2 = \theta_3 \times \theta_1$. We obtain ω by simply expanding the left-hand side of equation (52) using the computed θ.

APPENDIX B: SIMILARITIES AND DIFFERENCES OF TYPE-I AND TYPE-II PDES

We discussed the type-I and type-II PDEs, their solutions and the stable methods in sections IV and V. Here we first show that the type-I direct depth solution and type-II direct depth solution are exactly the same. Then we describe how the stable type-I and stable type-II methods can differ in their final reconstructions.

Equivalence of type-I and type-II solutions: The first non-holonomic solution of type-I PDE α and type-II PDE φ_z are given by equation (20) and equation (55) respectively. The type-I and type-II PDEs are obtained using the same deformation and reprojection constraints. With the equivalence of type-I and type-II PDEs, the non-holonomic solutions are directly related by the change in variable rewritten below:

$$\varphi_z = \frac{\alpha}{\nu} \quad \text{and} \quad \nu = \sqrt{1 + \frac{\eta^\top \eta}{f^2}}. \tag{iv}$$

The quantity ν involved in the change of variable is a known quantity and thus it proves that the direct depth reconstructions using type-I and type-II PDEs are the same.

In order to prove the equivalence of the second non-holonomic solutions, we consider the case of flat template, where we have $\mathsf{J}_\phi = \mathsf{J}_\varphi$. The result can be extended to non-flat templates by using the locally isometric flattening. We differentiate equation (iv) on both sides that gives us:

$$\mathsf{J}_\alpha = \nu \mathsf{J}_{\varphi_z} + \varphi_z \mathsf{J}_\nu. \tag{v}$$

Furthermore we already have the relation between J_{φ_z} and J_φ from equation (13) in section IV-A, rewritten below:

$$\mathsf{J}_\varphi = \mathbf{M}\tilde{\eta}\mathsf{J}_{\varphi_z} + \Phi \mathsf{J}_{\tilde{\eta}}. \tag{vi}$$

Thus substituting the value of J_{φ_z} from equation (v) in equation (vi), we have:

$$\mathsf{J}_\varphi = \mathbf{M}\tilde{\eta}\frac{(\mathsf{J}_\alpha - \varphi_z \mathsf{J}_\nu)}{\nu} + \Phi \mathsf{J}_{\tilde{\eta}}. \tag{vii}$$

In other words, one can go from the type-I set of solutions for $(\alpha, \mathsf{J}_\alpha)$ to type-II set of solutions for $(\varphi_z, \mathsf{J}_\varphi)$ using equations (iv) and (vii). Similarly to go from type-II set of solutions to type-I set we note the fact that J_{φ_z} is given by the third row of J_φ and thus obtain $(\alpha, \mathsf{J}_\alpha)$ by using equations (iv) and (v).

Difference in stable type-I and stable type-II reconstructions: As shown above, the non-holonomic solutions of the type-I and type-II PDES are equivalent up to some change in variable. Nevertheless, the second non-holonomic solutions measure different quantities. This naturally leads to differences in the subsequent steps: the resolution of the two-fold ambiguity and the numerical integration. The second non-holonomic solution of type-I PDE β or J_α has a sign ambiguity, whereas that of type-II PDE: J_ϕ or $\mathbf{n}$ has a two-fold ambiguity. The resolution of ambiguities in these two cases can produce outcomes that influence the further steps differently. More importantly, the numerical integration involved in the steps differ as the integrand in stable type-I method β is a 2-vector while the integrand in stable type-II method $\mathbf{n}$ is a vector describing the shape normal, *i.e.* a 3-vector. Noise can influence the numerical integration in the different spaces differently. The numerical integration is done by using LLS with the bending energy of BBS. It enforces smoothness on the integrand's space. Such smoothness is better suited on the space of surface normals than on the radial depth gradient where the norm of the image coordinates is involved. Thus the small but consistent difference seen in the reconstructions is originate from the influence of noise in the sequence of steps after the non-holonomic solutions are found and not from the stable PDE solutions themselves.

APPENDIX C: RECONSTRUCTION IN THE PRESENCE OF OUTLIERS

We use a sugar bottle with a non-developable 3D template to test how outliers affect the reconstruction and how they can be dealt with. The bottle is used to create a deformation which we use to test the effect of outliers on our best performing method, *i.e.* **typeII**-P. We take the images with the same camera focal length as the Can dataset shown in figure I. We combine SIFT and KAZE features to obtain the template image to input image warp. Outliers appear naturally in the matching due to deformation and repeated texture. We show that with outliers, SfT will fail if a standard

BBS registration [Dierckx, 1993] is used, while on the other hand using an automatic outlier removal method [Pilet et al., 2008; Pizarro and Bartoli, 2012] followed by a robust registration can suffice to have a good reconstruction.

Registration

Let $\{\mathbf{p}_i \in \mathbb{R}^2\}$ be the set of corresponding points in the template image and $\{\mathbf{q}_i \in \mathbb{R}^2\}$ be the set of corresponding points in the input image obtained from feature point matching, with $i = 1 \dots N$. We define the BBS registration function going from the template image to the input image as:

$$\Psi(\mathbf{p}_i, \mathbf{l}) = \mathbf{q}_i \tag{viii}$$

where $\mathbf{l} \in \mathbb{R}^{2n_c}$ is the parametrization vector of the warp function Ψ and n_c is the number of the control centers. The warp is estimated from the correspondence points by minimizing the following least-squares problem:

$$\bar{\mathbf{l}} = \underset{\mathbf{l}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\Psi(\mathbf{p}_i, \mathbf{l}) - \mathbf{q}_i\|_2. \tag{ix}$$

For BBS registration the function Ψ is a linear function given by $\Psi(\mathbf{p}_i, \mathbf{l}) = \mathbf{A}\mathbf{l}$ where $\mathbf{A}$ is a matrix of basis vectors constructed from the set of points $\{\mathbf{p}_i \in \mathbb{R}^2\}$ by imposing smoothness and is a known matrix. Thus equation (ix) is in fact a LLS problem for a standard BBS registration. However the presence of outliers in the correspondence points means one cannot correctly estimate $\mathbf{l}$ with the LLS problem of equation (ix). We can remove a large number of outliers using existing methods such as [Pilet et al., 2008; Pizarro and Bartoli, 2012]. This allows us to compute $\mathbf{l}$ using the remaining set of correspondences. However because these methods are not perfect, a small number of outliers may remain. To tackle this, we first employ the outlier rejection method [Pizarro and Bartoli, 2012] and fit $\mathbf{l}$ with a robust cost function, otherwise known as an M-estimator. Specifically we use the L1 M-estimator. Thus the minimization problem of equation (ix) is modified as follows:

$$\hat{\mathbf{l}} = \underset{\mathbf{l}}{\operatorname{argmin}} \sum_{i=1}^{N} \|\Psi(\mathbf{p}_i, \mathbf{l}) - \mathbf{q}_i\|_1. \tag{x}$$

We choose the L1 M-estimator because the resulting problem is convex and can be solved using an off-the-shelf solver such as the L1-magic [Candes and Tao, 2005]. It is fast to optimize and contains no addtional free parameter. However, it is a non-redescending M-estimator and can only handle a small percentage of outliers. When there is a higher percentage of outliers even after outlier rejection, one can use redescending M-estimators such as Tukey's bisquare M-estimator.

In the example, 174 *out of* 1617 *points* are actual outliers. We remove a set of 172 points with the outlier rejection method [Pizarro and Bartoli, 2012] and perform the robust registration. After the outlier rejection we estimate 44 *out of* 1445 *points* to be actual outliers. For better visualization we select one out of every 40 initial matches and show them along with the detected inlier matches in figure I. We also show one out of every 5 detected outliers in the bottom row of figure I.

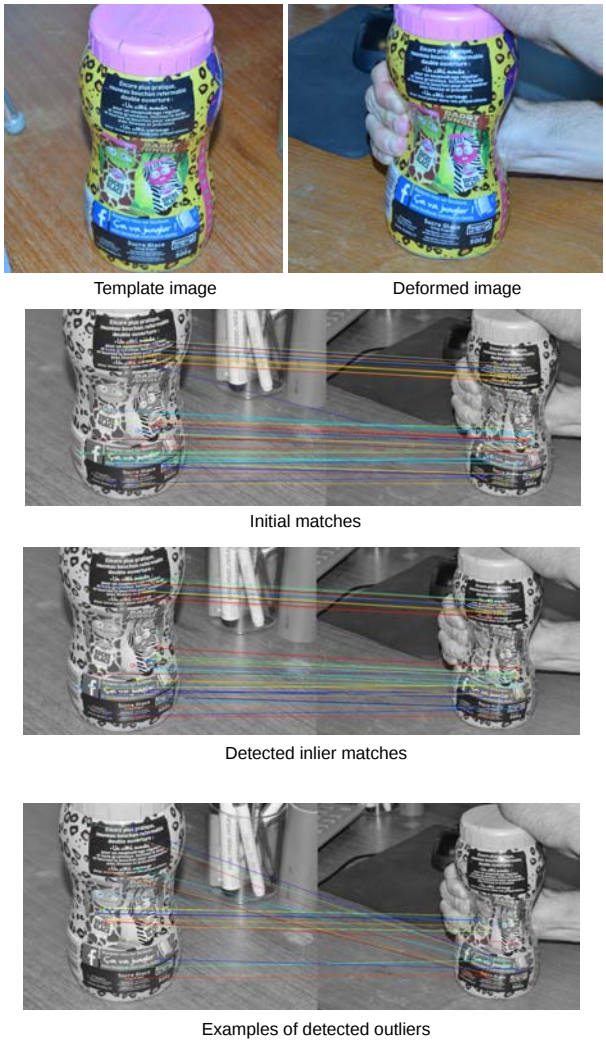


Fig. I: Cropped images and point matches with outliers

Reconstruction

Despite outlier removal and robust registration, some inaccuracies persist naturally in the final registration. The reconstruction step needs to be robust to noisy correspondences to be able to produce a good reconstruction in such scenario as mentioned in property *a)* in section I. We show the stable type-II texture mapped reconstructions with and without the outlier rejection [Pizarro and Bartoli, 2012] and the robust registration steps in figure II. The example proves the point that handling outliers in the registration step is enough to produce good reconstructions with the proposed method even in difficult cases.

REFERENCES

E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

P. Dierckx. *Curve and Surface Fitting with Splines*. Oxford University Press, Inc., 1993. ISBN 0-19-853441-8.

$e_d = 16.39\ mm$
$e_s = 41.27°$

$e_d = 2.70\ mm$
$e_s = 4.31°$

(a) (b) (c)

Fig. II: Results: (a) ground truth (b) **typeII**-P reconstruction with standard registration (c) **typeII**-P reconstruction with robust registration.

J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision*, 76(2):109–122, 2008.

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