Stable Template-Based Isometric 3D Reconstruction in All Imaging Conditions by Linear Least-Squares

Ajad Chhatkuli, Daniel Pizarro and Adrien Bartoli
ALCoV-ISIT, UMR 6284 CNRS / Université d'Auvergne, Clermont-Ferrand, France
`http://isit.u-clermont1.fr`

Abstract

It has been recently shown that reconstructing an isometric surface from a single 2D input image matched to a 3D template was a well-posed problem. This however does not tell us how reconstruction algorithms will behave in practical conditions, where the amount of perspective is generally small and the projection thus behaves like weak-perspective or orthography. We here bring answers to what is theoretically recoverable in such imaging conditions, and explain why existing convex numerical solutions and analytical solutions to 3D reconstruction may be unstable. We then propose a new algorithm which works under all imaging conditions, from strong to loose perspective. We empirically show that the gain in stability is tremendous, bringing our results close to the iterative minimization of a statistically-optimal cost. Our algorithm has a low complexity, is simple and uses only one round of linear least-squares.

1. Introduction

An important problem in computer vision is to infer a deformed 3D surface from its 2D projection in an image and a known 3D template. We here refer to this problem as Shape-from-Template (SfT). This is challenging due to the high variability and complexity of deformations in natural objects. Given template to input image registration, SfT has been studied for a variety of constraints including learnt shape bases [9, 14, 8], temporal smoothness [15], isometric deformation [3, 4, 7, 12], conformal deformation [4] and linear elastic deformation [1, 10].

We here specifically focus on isometric deformations which preserve surface geodesic distances. Isometric deformations apply to a wide variety of real surfaces and isometric SfT has been extensively studied [4, 3, 12, 14]. Existing approaches can be divided into three main groups: *i)* convex numerical optimization approaches [12, 14, 7], *ii)* analytical solutions [4, 3, 13] and *iii)* non-convex numerical

approaches [7]. The last group includes statistically optimal methods which need to be reliably initialized by a method from *i)* or *ii)* and are computationally expensive. Convex methods in *i)* are generally costly and may fail because of the relaxation they use of the non-convex isometry. Analytical solutions in *ii)* exploit the redundancy of isometric constraints to achieve local solutions. All existing methods in *i)* and *ii)* are in practice far from obtaining optimal results. There is thus a need for a method that gives fast and near-optimal solutions without requiring an initialization.

We prove that in orthographic conditions, depth is not directly recoverable, while depth-gradient is. In other words, a depth estimate is directly affected by the amount of perspective in the image. Clearly, orthography is never perfectly reached in practice. However, this result tells us that trying to estimate depth directly in SfT may be unstable. We show that indeed, all existing methods in *i)* and *ii)* are intrinsically unstable. Unfortunately, it is difficult to change the convex relaxations used in *i)*. We thus exploit a novel way of using the principle underlying methods in *ii)*, which use the non-holonomic depth solution to a system of nonlinear PDEs. We show that this solution is unstable under close to orthographic conditions, which tend to happen frequently in practice due to the local nature of these methods. We propose to use the two-fold non-holonomic solution to depth-gradient, which has not yet been exploited. Because we have shown that this solution is always stable, we can expect our method to give stable results. By using a simple disambiguation rule exploiting depth, stable even in close to orthographic conditions, and a linear least-squares integration step, we obtain a fast and simple SfT method. Our thorough experimental evaluation shows that its results are extremely close to a statistically optimal method in *iii)*, and that it largely outperforms existing methods in *i)* and *ii)*.

We review state-of-the-art in §2. We present our mathematical modeling and notation in §3. We derive the system of non-linear PDEs describing SfT and its solutions in §4. We study the solution stability with respect to projection conditions in §5. We present our stable solution in §6. We

show experimental results in §7 and conclude in §8.

Notation. We use bold for vectors (*e.g.* $\mathbf{p}$) and matrices (*e.g.* $\mathbf{A}$) and italics for scalars. We use Greek letters for functions (*e.g.* ψ). We use the operator J_φ to write the function giving the Jacobian matrix of φ. We use $\operatorname{diag}(\mathbf{v})$ to define a matrix whose diagonal is $\mathbf{v}$. 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 $i < j$.

2. Previous Work

Most methods in *i)* relax the non-convex isometric constraints to inextensibility. They then use the so-called Maximum Depth Heuristic (MDH) [12, 14]. The idea is to maximize the surface depth so that the Euclidean distance between every pair of points is upper bounded by its geodesic distance, known from the template. MDH methods are convex and fail if perspective is not strong. Methods in *ii)* use a system of nonlinear PDEs. Very recently [4, 3] found the analytical solution of the system for perspective and affine cameras. Analytical solutions form a powerful tool to study SfT, allowing one to prove the existence of solutions. However, those solutions are approximations when the registration warp contains noise [13] and do not ensure the surface to be exactly isometric [4]. Methods in *iii)* optimize a statistically optimal cost, that includes deformation and reprojection constraints. Those methods are non-convex and rely on iterative optimization [7] such as Levenberg-Marquardt.

This paper studies the properties of the solutions in *ii)* and finds that some of the algebraic solutions that are usually discarded in *ii)* (*i.e.* derivatives of depth or surface normals) can be the key to obtaining more accurate shapes, still keeping the low complexity of the analytical solutions with respect to optimization approaches in *i)* and *iii)*.

3. Modeling

3.1. Geometric Modeling

Figure 1 shows the modeling of SfT we use [4]. The template is a 2D domain $\Omega \subset \mathbb{R}^2$. Image registration is represented by a warp $\eta : \Omega \to \mathbb{R}^2$ from Ω to the input image $\mathcal{I}$. $\Delta : \Omega \to \mathcal{T}$ is invertible and parametrizes the 3D template surface $\mathcal{T} \subset \mathbb{R}^3$ from Ω. The template surface is deformed isometrically by $\psi : \mathcal{T} \to \mathbb{R}^3$. The deformed surface $\mathcal{S}$ is parametrized by the unknown embedding function $\varphi : \Omega \to \mathbb{R}^3$. The deformed surface is projected into $\mathcal{I}$ by the known camera $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$.

Our goal is to solve for SfT, represented by ψ. In practice we work with the embedding φ. This is equivalent since $\varphi = \psi \circ \Delta$. We solve for φ from the known functions Δ, η and Π, and the fact that the surface deforms isometrically.
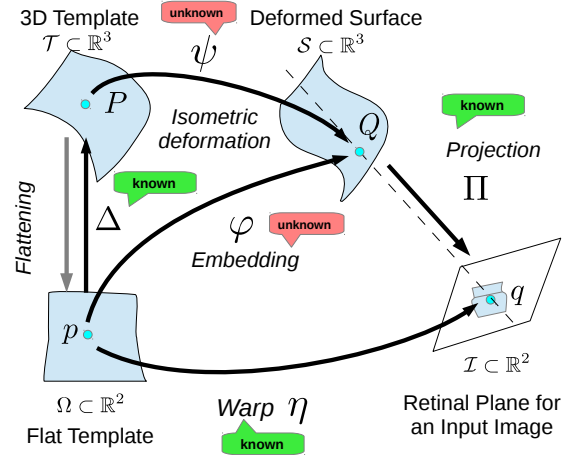

Figure 1. Geometric modeling of Shape-from-Template.

We divide the constraints on φ in reprojection constraints and deformation constraints.

3.2. Deformation Constraints

We start with the equation for the embedding $\varphi = \psi \circ \Delta$, whose differentiation leads to:

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

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:

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

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

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

3.3. Reprojection Constraints

The reprojection constraint $\eta = \Pi \circ \varphi$ 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. Let $f > 0$ be the camera's focal length. We use φ_z as the depth function, where $\varphi = (\varphi_x, \varphi_y, \varphi_z)^\top$.

3.3.1 The Perspective Camera

The perspective projection Π^{P} yields:

$$\eta = \Pi^{\mathrm{P}} \circ \varphi = \left(f \frac{\varphi_x}{\varphi_z}, \ f \frac{\varphi_y}{\varphi_z} \right)^{\top}. \qquad (5)$$

From equation (5) we use φ_z to parameterize the embedding φ, defining the following back-projection equation:

$$\varphi = \Phi^{\mathrm{P}} \tilde{\eta} \quad \text{with} \quad \Phi^{\mathrm{P}} = \mathrm{diag}\left(\frac{\varphi_z}{f}, \frac{\varphi_z}{f}, \varphi_z \right), \qquad (6)$$

and $\tilde{\eta}^{\top} = \left(\eta^{\top}, 1 \right)$.

3.3.2 Global and Local Weak-Perspective Cameras

The global weak-perspective camera approximates the perspective camera. It first projects the scene orthographically onto a fronto-parallel plane placed at the scene's average depth and then scales it. The local weak-perspective instantiates a weak-perspective camera at each point [5]. This gives the same projection as the perspective camera but simplifies the depth-gradient. It is a non-analytical model. Whether global or local the weak-perspective projection Π^{WP} yields:

$$\eta = \Pi^{\mathrm{WP}} \circ \varphi = \left(f \frac{\varphi_x}{\zeta}, \ f \frac{\varphi_y}{\zeta} \right)^{\top}. \qquad (7)$$

In the global weak-perspective model ζ is a constant function. In the local weak-perspective model ζ is different at each point, while preserving the property that $\mathsf{J}_\zeta = \mathbf{0}_{1 \times 2}$. The back-projection equation with both local and global weak-perspective is:

$$\varphi = \Phi^{\mathrm{WP}} \tilde{\eta} \quad \text{with} \quad \Phi^{\mathrm{WP}} = \mathrm{diag}\left(\frac{\zeta}{f}, \frac{\zeta}{f}, \varphi_z \right). \qquad (8)$$

4. PDEs and Non-Holonomic Solutions

From equation (6) or (8), the solution of SfT is to find the depth function φ_z, and thus φ, so that the deformation constraints (4) are met.

4.1. General PDEs

We derive a non-linear system of PDEs that holds for both perspective and weak-perspective projection (6) and (8) and deformation constraints (4). We first differentiate equation (6) and (8) to get J_φ, setting $\Phi \in \{\Phi^{\mathrm{P}}, \Phi^{\mathrm{WP}}\}$:

$$\mathsf{J}_\varphi = \mathbf{M}\tilde{\eta}\mathsf{J}_{\varphi_z} + \Phi \mathsf{J}_{\tilde{\eta}}, \qquad (9)$$

where:

$$\mathbf{M} = \frac{\partial \Phi}{\partial \varphi_z} = \begin{cases} \mathbf{M}^{\mathrm{P}} = \mathrm{diag}(\frac{1}{f}, \frac{1}{f}, 1) & \text{perspective} \\\\ \mathbf{M}^{\mathrm{WP}} = \mathrm{diag}(0, 0, 1) & \text{weak-perspective.} \end{cases} \qquad (10)$$

We introduce equation (9) in the isometric constraint (4) to get the following system of non-linear PDEs:

$$\mathsf{J}_{\varphi_z}^{\top} \tilde{\eta}^{\top} \mathbf{M}^2 \tilde{\eta} \mathsf{J}_{\varphi_z} + \mathsf{J}_{\varphi_z}^{\top} \tilde{\eta}^{\top} \mathbf{M}\Phi \mathsf{J}_{\tilde{\eta}}$$
$$+ \mathsf{J}_{\tilde{\eta}}^{\top} \Phi \mathbf{M}\tilde{\eta} \mathsf{J}_{\varphi_z} + \mathsf{J}_{\tilde{\eta}}^{\top} \Phi^2 \mathsf{J}_{\tilde{\eta}} = \mathsf{J}_\Delta^{\top} \mathsf{J}_\Delta. \qquad (11)$$

System (11) models SfT in terms of φ_z and J_{φ_z} for perspective and weak-perspective projections. Assuming J_{φ_z} and φ_z are independent variables, the algebraic solutions of system (11) are called non-holonomic solutions. We denote them as $\bar{\varphi}_z$ and $\bar{\kappa}$.

Non-holonomic solutions play an important role in the presence of noise or errors in the warp η. Despite the fact that system (11) 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 (11) is in fact an overdetermined system of PDEs with no general (*i.e.* holonomic) solutions.

We briefly present non-holonomic solutions of the system and discuss their properties for each projection model.

4.2. Perspective Solutions

The system of PDEs (11) is specialized to perspective projection by choosing Φ^{P} from equation (6) and $\mathbf{M}^{\mathrm{P}}$ from equation (10):

$$\left(1 + \frac{\eta^{\top}\eta}{f^2} \right) \mathsf{J}_{\varphi_z}^{\top} \mathsf{J}_{\varphi_z} + \frac{\varphi_z}{f^2} (\mathsf{J}_{\varphi_z}^{\top} \eta^{\top} \mathsf{J}_\eta + \mathsf{J}_\eta^{\top} \eta \mathsf{J}_{\varphi_z})$$
$$+ \frac{\varphi_z^2}{f^2} \mathsf{J}_\eta^{\top} \mathsf{J}_\eta = \mathsf{J}_\Delta^{\top} \mathsf{J}_\Delta. \qquad (12)$$

We simplify this system by changing variables with:

$$\alpha = \varphi_z \nu \quad \text{and} \quad \nu = \sqrt{1 + \frac{\eta^{\top}\eta}{f^2}}, \qquad (13)$$

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 system of PDEs in α and J_α:

$$\mathsf{J}_\alpha^{\top} \mathsf{J}_\alpha + \alpha^2 \gamma = \mathsf{J}_\Delta^{\top} \mathsf{J}_\Delta, \qquad (14)$$

where:

$$\gamma = \frac{1}{\nu^2 f^2} \left(\mathsf{J}_\eta^{\top} \mathsf{J}_\eta - \frac{1}{\nu^2 f^4} \mathsf{J}_\eta^{\top} \eta \eta^{\top} \mathsf{J}_\eta \right). \qquad (15)$$

Following [4, 3] we can always find a single algebraic solution of system (14). We denote the non-holonomic solutions of α and J_α as $\bar{\alpha}$ and $\bar{\beta}$:

$$\bar{\alpha} = \sqrt{\lambda_2 \left(\mathsf{J}_\Delta^{\top} \mathsf{J}_\Delta \gamma^{-1} \right)} \qquad (16)$$

$$\bar{\beta} = \pm \sqrt{\lambda_1(\Upsilon)} \mathbf{v}_1(\Upsilon), \qquad (17)$$

where:

$$\Upsilon = J_\Delta^\top J_\Delta - \lambda_2 \left(J_\Delta^\top J_\Delta \gamma^{-1} \right) \gamma. \qquad (18)$$

We may recover $\bar\varphi_z$ from equation (16) followed by the change of variable (13). Instead of $\bar\kappa$, we recover $\bar\beta$ from equation (17), then integrate the solution followed by the change of variable (13) to obtain $\hat\varphi_z$. $\bar\beta$ alone retains the information in the non-holonomic solution $\bar\kappa$ up to an unknown sign and scale change [4].

4.3. Weak-Perspective Solutions

The general PDE for global and local weak-perspective cameras is found by choosing Φ^{WP} from equation (8) and M^{WP} from equation (10):

$$J_{\varphi_z}^\top J_{\varphi_z} + \frac{\zeta^2}{f^2} J_\eta^\top J_\eta = J_\Delta^\top J_\Delta, \qquad (19)$$

where we set $\zeta = \varphi_z$ in the local weak-perspective model as ζ gives the 'average' depth at a differential level. In the global weak-perspective model, $\zeta = z_a$ is constant. [3] gives a method to find z_a by integration over the whole surface domain and [13] studies the multiple solutions for $\bar\kappa$.

System (19) has exactly the same structure as system (14), the simplified PDE system for perspective projection. We denote the non-holonomic solutions of the system as $\bar\varphi_z$ and $\bar\kappa$. In the local weak-perspective model we obtain $\bar\varphi_z$ and $\bar\kappa$ by simply using equation (16) and (17) with $\gamma = f^{-2} J_\eta^\top J_\eta$.

4.4. Obtaining the Embedding

[4, 3] use $\bar\varphi_z$ directly to get the embedding φ through equation (6) and (8), neglecting the information contained in $\bar\beta$ and consequently $\bar\kappa$. At first glance that seems to be sensible as $\bar\beta$ is known only up to sign and requires integration to get depth. We show however, that $\bar\beta$ provides more accurate reconstructions of φ than $\bar\varphi_z$.

5. Stability of Non-Holonomic Solutions

5.1. Main Results

We prove two important results regarding the stability of the non-holonomic solutions of the general PDEs (11):

- **Result 1):** *The non-holonomic solution for depth $\bar\varphi_z$ is weakly constrained when the projection conditions tend to orthographic.*

- **Result 2):** *The non-holonomic solution for the depth gradient $\bar\kappa$ is well-constrained in all projection conditions.*

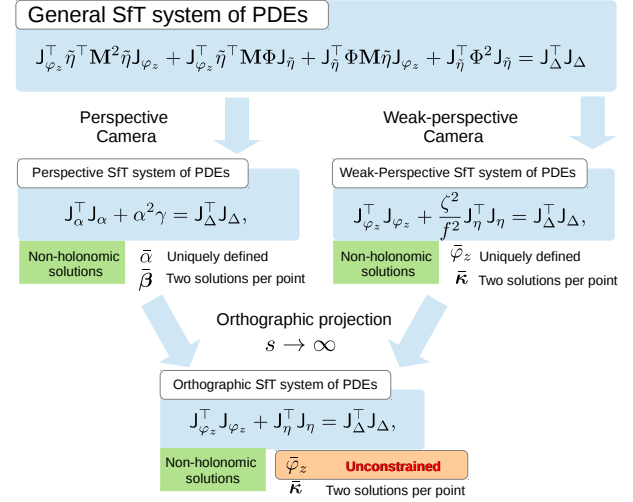Figure (2) shows a general diagram showing the effect of projection conditions on SfT.

Figure 2. SfT solutions for different projection models and amount of perspective.

To prove these results, we first define a projection function Π_s depending on a parameter s that allows us to continuously select the amount of perspective:

$$\Pi_s(\mathbf{Q}) = \frac{(s+1)f}{Q_z + sf} \begin{pmatrix} Q_x & Q_y \end{pmatrix}^\top. \qquad (20)$$

Equation (20) gives an orthographic projection when $s \to \infty$:

$$\lim_{s \to \infty} \Pi_s(\mathbf{Q}) = \begin{pmatrix} Q_x & Q_y \end{pmatrix}^\top. \qquad (21)$$

The weak-perspective approximation of Π_s is:

$$\Pi_s^{WP}(\mathbf{Q}) = \frac{(s+1)f}{\zeta + sf} \begin{pmatrix} Q_x & Q_y \end{pmatrix}^\top. \qquad (22)$$

5.2. Proof for the Perspective Camera

We first integrate the projection model Π_s in the general system of PDEs (11) by simply re-defining the backprojection matrix Φ for perspective projection as:

$$\Phi_s = \text{diag}\left(\frac{\varphi_z + sf}{(s+1)f}, \frac{\varphi_z + sf}{(s+1)f}, \varphi_z + sf \right). \qquad (23)$$

Introducing Φ_s in the general system of PDEs (11) we obtain:

$$\left(1 + \frac{\eta^\top \eta}{((s+1)f)^2} \right) J_{\varphi_z}^\top J_{\varphi_z} + \frac{\varphi_z + sf}{((s+1)f)^2} (J_{\varphi_z}^\top \eta^\top J_\eta) +$$

$$+ \frac{\varphi_z + sf}{((s+1)f)^2} (J_\eta^\top \eta J_{\varphi_z}) + \frac{(\varphi_z + sf)^2}{((s+1)f)^2} J_\eta^\top J_\eta = J_\Delta^\top J_\Delta. \quad (24)$$

By taking the limit as $s \to \infty$ on both sides of equation (24) we find the following system:

$$\mathsf{J}_{\varphi_z}^\top \mathsf{J}_{\varphi_z} + \mathsf{J}_\eta^\top \mathsf{J}_\eta = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta, \qquad (25)$$

which represents the general system of PDEs for orthographic projection [13]. In equation (25) the depth variable φ_z vanishes, which means that with orthographic projection depth is not anymore constrained.

Proof of result 1): When s is a large number, the solution of $\bar{\alpha}$ in equation (16) is not well conditioned. $\bar{\alpha}^2$ depends on the eigenvalues of matrix $\mathsf{J}_\Delta^\top \mathsf{J}_\Delta \gamma^{-1}$. We write γ from equation (15) as a function of s:

$$\gamma_s = \frac{1}{\nu_s^2 \left((s+1)f\right)^2} \left(\mathsf{J}_\eta^\top \mathsf{J}_\eta - \frac{1}{\nu_s^2 \left((s+1)f\right)^4} \mathsf{J}_\eta^\top \eta \eta^\top \mathsf{J}_\eta \right), \qquad (26)$$

with:

$$\nu_s = \sqrt{1 + \frac{\eta^\top \eta}{\left((s+1)f\right)^2}}. \qquad (27)$$

Taking the limit of equation (26) 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.

Proof of result 2): By applying the rank-1 constraint to equation (25), the solution of $\bar{\kappa}$ is simply given by:

$$\bar{\kappa} = \pm\lambda_1 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \mathsf{J}_\eta^\top \mathsf{J}_\eta \right) \mathbf{v}_1 \left(\mathsf{J}_\Delta^\top \mathsf{J}_\Delta - \mathsf{J}_\eta^\top \mathsf{J}_\eta \right). \qquad (28)$$

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

5.3. Proof for Weak-Perspective Cameras

By taking the weak-perspective approximation of the projection model Π_s^{WP} and plugging it into equation (19) we reach the following system:

$$\mathsf{J}_{\varphi_z}^\top \mathsf{J}_{\varphi_z} + \left(\frac{\zeta + sf}{(s+1)f} \right)^2 \mathsf{J}_\eta^\top \mathsf{J}_\eta = \mathsf{J}_\Delta^\top \mathsf{J}_\Delta. \qquad (29)$$

Again by taking the limit $s \to \infty$ on both sides of equation (29), we reach the system (25) of orthographic projection. If there is no perspective effect and the camera is orthographic we cannot get the average depth of the scene as it vanishes from the equations.

The proof of result 1) follows in the same way as with the perspective camera by using $\gamma = \frac{1}{((s+1)f)^2} \mathsf{J}_\eta^\top \mathsf{J}_\eta$. For result 2) the solution of $\bar{\kappa}$ in equation (28) is identical with the weak-perspective model when $s \to \infty$.
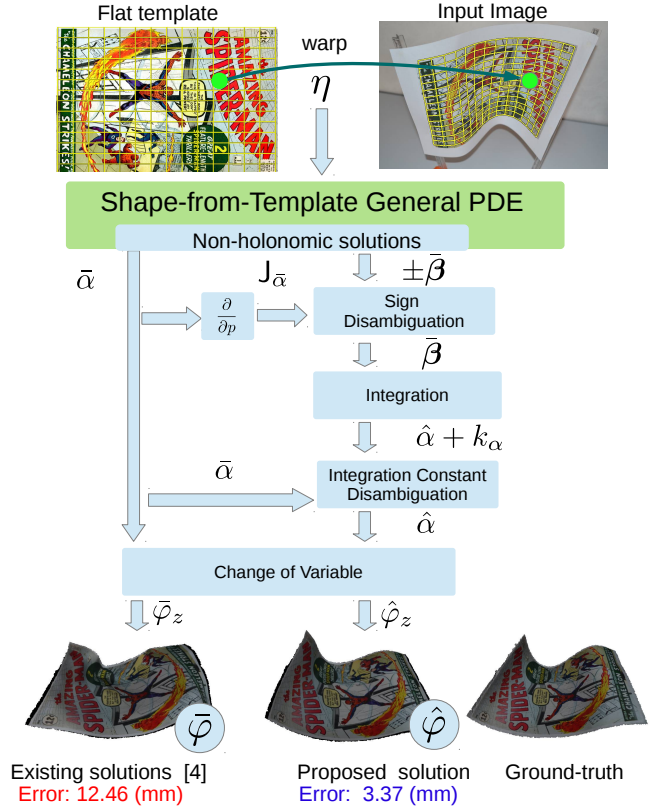

Figure 3. Proposed SfT method using the non-holonomic solution to depth-gradient. This method is stable under all imaging conditions.

6. SfT from Depth-Gradient

Because depth is locally unstable, we propose to use $\bar{\beta}$ to solve SfT. In order to get depth $\hat{\varphi}_z$ from $\bar{\beta}$ we need to solve the following three problems: *i)* sign disambiguation for $\bar{\beta}$, *ii)* integration of $\bar{\beta}$ and *iii)* the arbitrary integration factor. Figure 3 shows the general diagram of our method.

6.1. Sign Disambiguation

According to equation (17), non-holonomic solutions for $\bar{\beta}$ and $\bar{\kappa}$ are known up to a local sign change. [4, 13] propose a few methods to disambiguate the sign, at least partially, 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 $\bar{\beta}$ or $\bar{\kappa}$ by using the non-holonomic solution to depth $\bar{\varphi}_z$.

In the perspective camera the process has three steps: **1)** We first differentiate $\bar{\alpha}$ to obtain $\mathsf{J}_{\bar{\alpha}}$. **2)** We discard those regions of the template where $\mathsf{J}_{\bar{\alpha}}$ differs substantially from

$\bar{\boldsymbol{\beta}}$. We use the angle between the two vectors as a metric:

$$\theta = \text{acos}\left(\frac{|\mathsf{J}_{\bar{\alpha}}\bar{\boldsymbol{\beta}}|}{\|\mathsf{J}_{\bar{\alpha}}\|\|\bar{\boldsymbol{\beta}}\|}\right). \tag{30}$$

3) We select the sign of $\bar{\boldsymbol{\beta}}$ so that the resulting vector is the closest to $\mathsf{J}_{\bar{\alpha}}$.

With the weak-perspective camera the three steps are almost identical, using $\bar{\varphi}_z$ to disambiguate $\bar{\boldsymbol{\kappa}}$. Here, with no change of variable required, $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\kappa}}$ are identical.

6.2. Numerical Integration

The non-holonomic solution $\bar{\boldsymbol{\beta}}$ 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 Thin-Plate Spline (TPS). With a TPS, or any other linear basis expansion model, we can integrate $\bar{\boldsymbol{\beta}}$ by means of linear least squares. The solution is defined up to an additive integration constant.

6.3. Integration Constant

After integration we obtain $\hat{\alpha}_z + k_z$, where k_z is an arbitrary integration constant. We propose to use $\bar{\alpha}$ to estimate k_z. Given a set of points $\{\mathbf{p}_i\}_{i=1}^{N}$ in the template domain we obtain samples from $\hat{\alpha} + k_z$ and $\bar{\alpha}$. We then obtain k_z by using the median of the differences between the samples.

7. Experimental Results

7.1. Compared Methods and Error Measurements

We use MATLAB to perform experiments with our methods[1] as well as with the compared methods. In all experiments we use two types of error measurements to quantify accuracy: mean depth error (average distance between the reconstructed and true surfaces) in mm and mean shape error (average difference between the reconstructed and true surface normals) in degrees. In all experiments we compare the following algorithms: **AnD** prefixes methods using directly the non-holonomic solution for depth $\bar{\varphi}_z$. **AnJ** is the method we propose in §6 using non-holonomic solutions for depth-gradient. We write **AnD-P** [4] and **AnJ-P** for the perspective camera model and **AnD-WP** [3] and **AnJ-WP** for the weak-perspective camera model. We denote as **ReD-P** and **ReJ-P** the methods **AnD-P** and **AnJ-P** followed by non-linear refinement using [7]. We consider those methods as giving the optimal results. We also compare with methods **Salz** [15] and **Perr** [12] based on the inextensible relaxation and the Maximum Depth Heuristic.

We use TPS for representing functions η, Δ and φ. We impose smoothness using the bending energy [6]. The following problems are all solved with linear least-squares:

a) to obtain η from a set of feature correspondences and smoothness priors [2]. b) to fit a scalar function (*e.g.* $\bar{\varphi}_z$) from a set of sample points, prior to computing its derivatives (*e.g.* $\mathsf{J}_{\bar{\varphi}_z}$) in closed form.

7.2. Synthetic Data

We use developable surfaces [11] to simulate 6 different isometric deformations of a flat template of size 640 px $\times$ 480 px. We generate synthetic images using a pin-hole camera with varying focal length. We use a single parameter s to define the focal length $f = (s + 1)500$ px. With $s = 0$ the camera has a focal length of 500 px. The scene depth is translated with s to keep the size of the object in the image invariant against f. We randomly generate N point correspondences between the template and the image and add Gaussian noise to their positions with standard deviation σ in px. We use $s = 0$, $\sigma = 1.0$ px and $N = 100$ as default values in all experiments. Figure 4 shows our results.

All experiments clearly show that **AnJ** methods are very accurate and close to their non-linear refinement **ReJ** in both depth and shape errors. Our proposal works consistently better than **AnD** methods that use the non-holonomic solution to depth. In the presence of very high correspondence noise ($\sigma = 3$), **AnJ-P** has a mean depth error of 20 mm and a mean shape error of about 9 degrees, which is even slightly better than **ReD**. With very low number of correspondences, we observe that our method has the least error after the statistically optimal method. Against increasing focal length, though **Salz** captures the shape better in the range between $s = 3$ and $s = 11$, it fails to estimate depth. In the range $f = 100$ px to $f = 1500$ px, which is also the working range of most cameras, **AnJ-P** has the best shape and depth errors after **ReJ** and **ReD**. We see no relevant differences between the camera models in the analytical methods when $s > 2$. Similar observations can be made for varying number of correspondences.

7.3. Real Data

The CVLAB Paper dataset. The CVLAB dataset [16] consists of 191 frames taken with about the same angle and focal length of a sheet of paper being deformed. The number of features detected in each frame of the sequence is around $N = 1300$. The performance of the different methods for each frame is plotted in figure 5.

The results with this dataset are very similar to those obtained on the synthetic data. In this case we can see that **Salz** needs a high number of features. This was also seen in the third column of figure 4 for the synthetic dataset experiment. We found that **Perr** has depth and shape errors outside the scale of the plots in the first row of figure 5, as the number of correspondences gets larger. Although this dataset presents a favorable scenario for methods like **Salz**

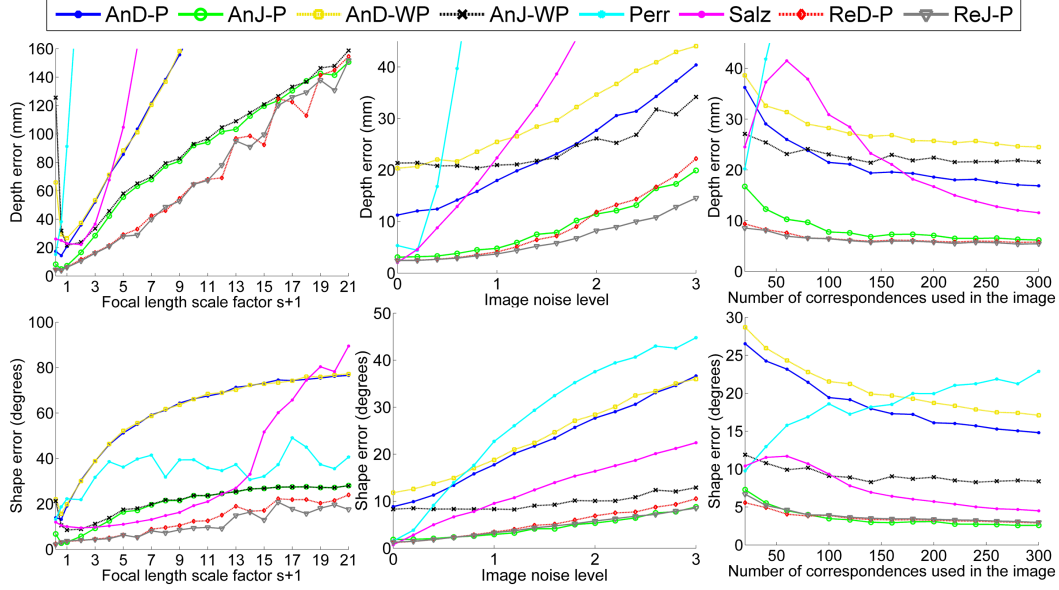[1]The codes are available at http://isit.u-clermont1.fr/~ab/Research/index.html.

Figure 4. Synthetic data experiments. We show the depth errors in the first row and the shape errors in the second row. In the first column we show the errors against the parameter s. In the middle column we show the influence of σ and in the last column the influence of N.
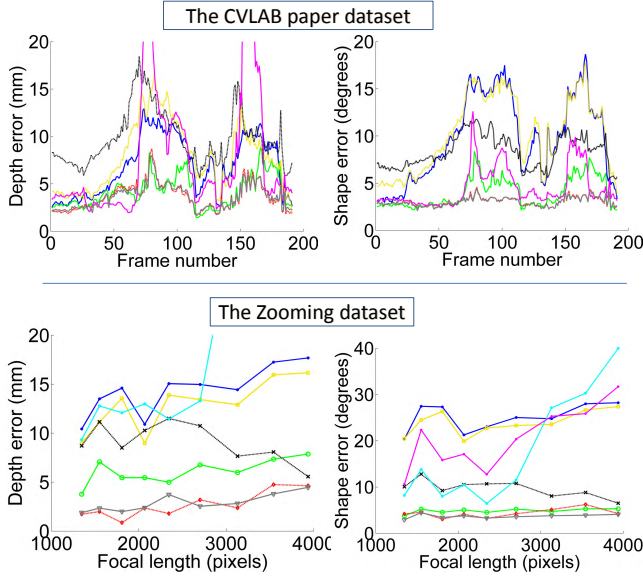

Figure 5. Plots for the CVLAB dataset and Zooming dataset (legend in figure 4).

and views. The focal length varies from 1300 px to 4000 px for an image size of 1728×1552 px. Each zoom level has between 7-10 images with different viewing angles. We computed the ground-truth from each view in camera coordinates using stereo. We computed the depth and shape errors for each image in the sequence. The second row in figure 5 shows the mean errors over different views against the focal length. To illustrate the reconstruction accuracy for each point, we show the reconstruction of all 8 methods, texture mapped with a colormap representing the shape error in figure 6. The illustration is presented for the zoom levels 1, 4 and 9. We denote as Mde and Msh, the depth and shape errors respectively. Our solution **AnJ-P** performs consistently well over all focal lengths, with a total mean depth error of about 6.08 mm compared to 77.74 mm for **Salz** and 2.63 mm for **ReD**. Furthermore, as the focal length goes over 4000 px, the camera models converge as expected.

8. Conclusions

We have shown that depth-gradient is always recoverable in SfT, and consequently so is relative depth. Non-convex statistically optimal algorithms are thus well-formulated. However, we have shown that existing initialization algorithms are unstable in non-strongly perspective imaging conditions. This includes both convex algorithms based on the maximum-depth heuristic and analytical solutions based on keeping the non-holonomic solution to depth.

We have proposed to keep the non-holonomic solution to depth-gradient which, contrarily to depth, is always sta-

we observe that our approach is still performing better in most of the frames for both shape and depth errors. The mean errors for the depth are: 6.9 mm for **AnD-P**, 4.18 mm for **AnJ-P**, 7.76 mm for **Salz** and 3.62 mm for **ReD-P**. As the images are very perspective, **AnJ-WP** does not perform well. Similar results also hold for the shape error.

The Zooming dataset. We propose a dataset which shows folded sheets of paper with different focal lengths
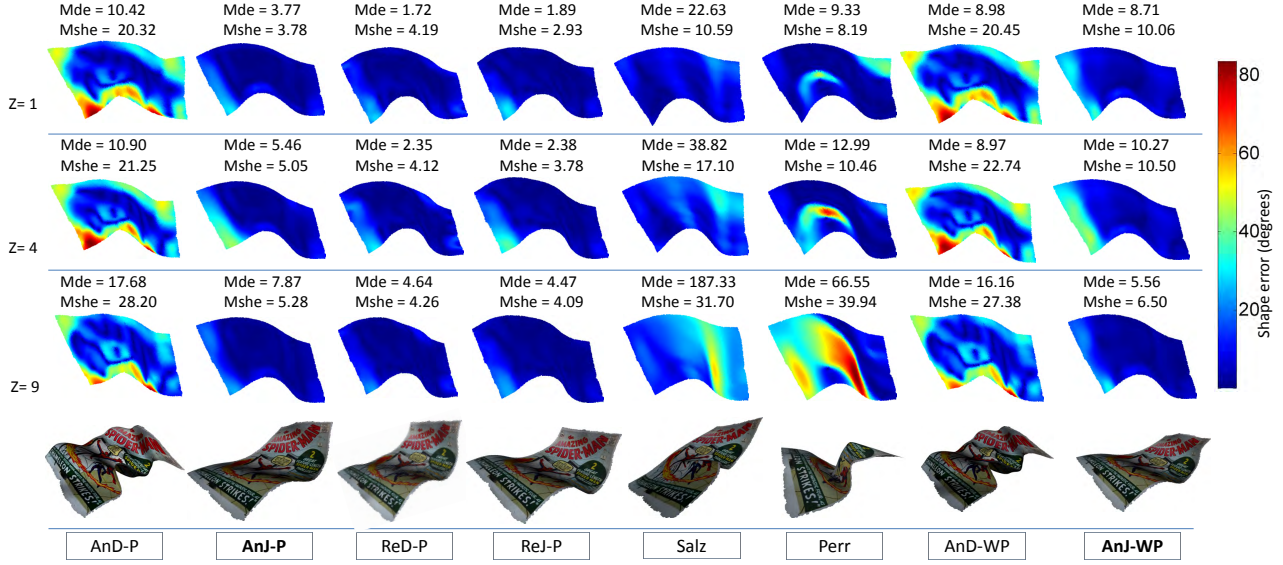
Figure 6. Representation of the shape error in 3 zoom levels of the Zooming dataset and the 3D shapes reconstructed for zoom level-9.

ble. Our algorithm is mostly analytical and requires only linear least-squares. It does not use perspective directly, but only to disambiguate the depth-gradient field and to find the integration constant. It is therefore fast, simple and stable. It outperforms state-of-the-art and gives results extremely close to statistically-optimal refinement.

References

[1] A. Agudo, B. Calvo, and J. Montiel. Finite element based sequential bayesian non-rigid structure from motion. In *CVPR*, 2012. 1

[2] A. Bartoli. Maximizing the predictivity of smooth deformable image warps through cross-validation. *Journal of Mathematical Imaging and Vision*, 31(2):133–145, 2008. 6

[3] A. Bartoli and T. Collins. Template-based isometric deformable 3D reconstruction with sampling-based focal length self-calibration. In *CVPR*, 2013. 1, 2, 3, 4, 6

[4] A. Bartoli, Y. Gerard, F. Chadebecq, and T. Collins. On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces. In *CVPR*, 2012. 1, 2, 3, 4, 5, 6

[5] A. Bartoli, D. Pizarro, and T. Collins. A robust analytical solution to isometric shape-from-template with focal length calibration. In *ICCV*, 2013. 3

[6] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. 6

[7] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres. Monocular template-based reconstruction of smooth and inextensible surfaces. In *ACCV*, 2010. 1, 2, 6

[8] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *CVPR*, 2012. 1

[9] A. Del Bue, X. Llad, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *CVPR*, 2006. 1

[10] A. Malti, R. Hartley, A. Bartoli, and J. Kim. Monocular template-based 3d reconstruction of extensible surfaces with local linear elasticity. In *CVPR*, 2013. 1

[11] 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. 6

[12] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *International Journal of Computer Vision*, 95(2):124–137, 2011. 1, 2, 6

[13] D. Pizarro, A. Bartoli, and T. Collins. Isowarp and conwarp: Warps that exactly comply with weak-perspective projection of deforming objects. In *BMVC*, 2013. 1, 2, 4, 5

[14] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(5):931–944, 2011. 1, 2

[15] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-D tracking. In *ICCV*, 2007. 1, 6

[16] A. Varol, M. Salzmann, P. Fua, and R. Urtasun. A constrained latent variable model. In *CVPR*, 2012. 6