A Robust Analytical Solution to Isometric Shape-from-Template with Focal Length Calibration

Adrien Bartoli, Daniel Pizarro and Toby Collins
ALCoV-ISIT, UMR 6284 CNRS / Université d'Auvergne, Clermont-Ferrand, France

Abstract

We study the uncalibrated isometric Shape-from-Template problem, that consists in estimating an isometric deformation from a template shape to an input image whose focal length is unknown.

Our method is the first that combines the following features: solving for both the 3D deformation and the camera's focal length, involving only local analytical solutions (there is no numerical optimization), being robust to mismatches, handling general surfaces and running extremely fast. This was achieved through two key steps. First, an 'uncalibrated' 3D deformation is computed thanks to a novel piecewise weak-perspective projection model. Second, the camera's focal length is estimated and enables upgrading the 3D deformation to metric. We use a variational framework, implemented using a smooth function basis and sampled local deformation models. The only degeneracy – which we easily detect– for focal length estimation is a flat and fronto-parallel surface.

Experimental results on simulated and real datasets show that our method achieves a 3D shape accuracy slightly below state of the art methods using a precalibrated or the true focal length, and a focal length accuracy slightly below static calibration methods.

1. Introduction

3D reconstruction from a single image and a template (a known 3D view of the surface) has been researched actively over the past decade. We here call this problem *Shape-from-Template* (*SfT*). Recovering the 3D deformation is equivalent to recovering the shape as seen in the input image. Solving SfT requires one to constrain the space of possible 3D deformations between the template and the unknown shape. An important instance of SfT is *IsoSfT*, where the 3D deformation is distance-preserving, in other words, an isometry. IsoSfT has been the most studied instance of SfT [2, 3, 4, 6, 8, 10] and was shown to generally admit a unique solution [2]. Importantly, most previous work assume known intrinsic camera calibration.

We are here interested in *C-IsoSfT*, the IsoSfT problem which takes an uncalibrated image as input and includes camera calibration as an unknown. We give a general framework and a detailled solution to the most important practical case where all camera parameters are known (the principal point, aspect ratio and skew) but the focal length. For most applications of SfT, being able to estimate the focal length online is the most important case since it allows one to zoom in and out while filming the deformable surface. More specifically, we contribute with the first robust analytical solution to recover 3D shape and the camera's focal length. We implemented our theory using putative keypoint correspondences as inputs. Our implementation discards erroneous correspondences and is entirely analytical in that it does not involve numerical optimization. This is important in two respects: ensuring that the solution is globally optimal and that it can be computed extremely fast.

There are two key differences between our framework and state of the art: *(i)* most current approaches require known camera parameters and *(ii)* most current approaches involve numerical optimization, except [2, 4]. Our analytical solution is based on a variational problem formulation with general template formulation. In a first step, we instantiate it with a novel projection model we call *Piecewise Weak-Perspective* (PWP). This allows us to derive an operator which locally maps an image warp to an uncalibrated solution to 3D shape. In a second step, we robustly solve for the focal length and upgrade 3D shape globally. Both steps involve analytical solutions and are extremely fast to compute. We believe that SfT is a local problem in nature. In other words, correspondences do not contribute aways from their local area of influence. There is however a trade-off between locality and stability. Our implementation handles this trade-off by constructing a multiple scale pool of local image warps.

Notation. We use greek for functions (*e.g.* η), italic latin for scalars (*e.g.* a), bold latin for vectors and matrices (*e.g.* $\mathbf{J}$) and double bars for domains (*e.g.* $\mathbb{R}$). The identity matrix is written $\mathbf{I}$. We define $\mathbb{S}^p$ as the space of symmetric positive matrices of size $(p \times p)$ and $\mathbb{O}$ as the space

of column-orthonormal matrices. We write $C^p(\mathbb{M}_1, \mathbb{M}_2)$ the space of p times continuously differentiable functions with domain $\mathbb{M}_1$ and codomain $\mathbb{M}_2$. We denote the largest/smallest eigenvalue functions as $\lambda_{1,2} \in C^1(\mathbb{S}^2, \mathbb{R})$, and define $\epsilon \in C^0(\mathbb{S}^2, \mathbb{R}^2)$ as giving the principal vector of a rank-1 matrix (*i.e.* $\epsilon(\mathbf{u}\mathbf{u}^\top) \stackrel{\text{def}}{=} \mathbf{u} \in \mathbb{R}^2$).

2. State of the Art

Existing SfT methods can be broadly classified into three categories: *(C1)* analytical solutions, *(C2)* convex optimization and *(C3)* nonconvex optimization.

(C1). Both closest works to ours lie in *(C1)*. They derive variational solutions to SfT using perspective [2] and orthographic [4] projection, but do not address camera calibration. None of [2, 4] copes with mismatches, and none lends itself to camera calibration (perspective projection does not allow one to factor out the focal length, and thus to compute an uncalibrated solution). A major advantage of variational solutions is that they may be extremely fast to compute.

(C2). Because IsoSfT is in nature nonconvex, methods in *(C2)* use a convex relaxation of the constraints. The most successful relaxation is the so-called inextensibility, which upper bounds the Euclidean distance between a pair of points by its true geodesic distance computed from the template [8]. With this relaxation, the cost is convex and leads to an SOCP. However, a term that prevents the surface from shrinking to the origin is required. This was implemented by the maximum-depth heuristic [3, 10] and penalized slack variables [6]. Mismatches may be handled by an annealing process [10, 6], though this makes the overall process nonconvex.

(C3). Methods in *(C3)* estimate a quasi-isometric deformation, which is a nonconvex constraint, while minimizing the reprojection error [3]. They may substantially improve an initial 3D shape estimate provided by a *(C1)* or *(C2)* algorithm [2].

Finally, a recent paper has also shown that the focal length could be calibrated in SfT [1]. The key idea is to sample a set of admissible focal lengths, solve SfT for each of them and keep the one minimizing some consistency measure.

In regards to state of the art, we present the first framework with the following features: *(i)* to solve SfT and camera focal length, *(ii)* to discard erroneous matches, *(iii)* to run extremely fast, *(iv)* to use a generic problem formulation independent of template parameterization, *(v)* to use a multi-scale set of image warps and *(vi)* without numerical optimization. This is achieved by first solving the problem locally to get an initial uncalibrated shape and then estimating the focal length. At the heart of the first step lies the Piecewise Weak-Perspective (PWP) projection model that we are presenting.

3. Generic C-IsoSfT Problem Formulations

We first give C-IsoSfT's 3D formulation, and then two other formulations based on the embedding function. All three formulations are equivalent. Each formulation has a data constraint called the *reprojection constraint* and a prior called the *deformation constraint*. Our goal is to arrive at a formulation which is solvable *locally*. This fundamental property holds for our second embedding-based 'point-normal' formulation. The problem setup is illustrated in figure 1. Our three formulations hold as well for IsoSfT.

3.1. Modeling and 3D Formulation

In the SfT problem, one has a 3D shape template $\mathcal{R} \subset \mathbb{R}^3$ which may be represented as a parametric surface with an embedding $\zeta \in C^2(\Omega, \mathbb{R}^3)$ from a parameterization space $\Omega \subset \mathbb{R}^2$. Given one image of the deformed 3D shape $\mathcal{S}$, the unknowns are *(i)* the 3D deformation $\Psi \in C^2(\mathcal{R}, \mathbb{R}^3)$ that brings $\mathcal{R}$ to $\mathcal{S}$, and *(ii)* the camera projection function $\Pi \in \mathcal{P}$. Here $\mathcal{P}$ is a space of 'intrinsic' camera projection functions, which we keep abstract for now. Our practical solution estimates solely the focal length which is the most important intrinsic of the pinhole camera. The extrinsic camera parameters are included in Ψ.

The data constraints in SfT are image matching constraints between the template and the input image. We model them as an image warp $\eta \in C^2(\Omega, \mathbb{R}^2)$, though they may also be represented by keypoint matches [9]. In particular, our implementation relies on local image warps, and η is thus not computed in practice.

Proposition 1 (3D Formulation of C-IsoSfT)

$$\underset{\Psi \in C^2(\mathcal{R}, \mathbb{R}^3), \Pi \in \mathcal{P}}{\text{find}} \begin{cases} \eta = \Pi \circ \Psi \circ \zeta & \text{(reprojection)} \quad (1) \\ \nabla_{\nabla_{\mathcal{R}} \Psi} \in C^1(\mathcal{R}, \mathbb{O}) & \text{(deformation)} \quad (2) \end{cases}$$

Proof of proposition 1. The reprojection constraint (1) is obvious by construction (see figure 1). The deformation constraint (2) means that the Jacobian matrix of Ψ in the tangent plane at any point of $\mathcal{R}$ has to be a column-orthonormal matrix. This must hold for Ψ to represent an isometric deformation of $\mathcal{R}$. In this equation, $\nabla_{\mathbf{D}}$ stands for the directional derivatives in the subspace $\mathbf{D}$.

3.2. Embedding-based Point-Tangent Formulation

We define the deformed surface embedding function $\varphi \in C^2(\Omega, \mathbb{R}^3)$ as:

$$\varphi \stackrel{\text{def}}{=} \Psi \circ \zeta. \quad (3)$$

We now give a formulation of C-IsoSfT that depends on the embedding φ (the 'point') and its partial derivatives (the 'tangent plane'). We define $\mathcal{T}$ as the operator that forms a function's metric tensor: $\mathcal{T}\varphi \stackrel{\text{def}}{=} \nabla\varphi^\top \nabla\varphi$.

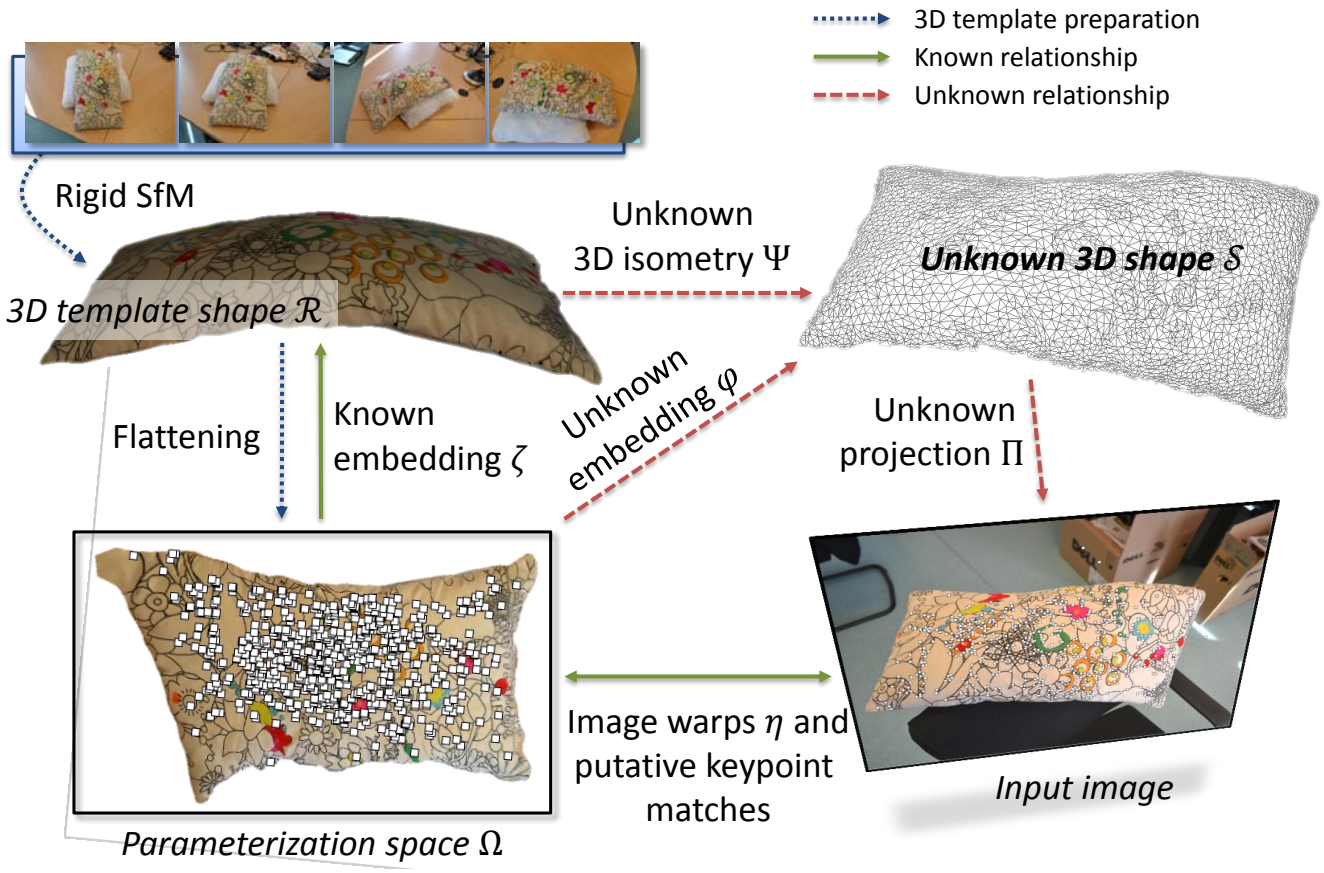


Figure 1. **General modeling of the C-IsoSfT problem.** The template shape may have an arbitrary topology and parameterization. A similar modeling was used in the literature [2] which studies IsoSfT with a known projection operator Π.

Proposition 2 (Point-Tangent Formulation of C-IsoSfT)

$$\underset{\varphi\in C^2(\Omega,\mathbb{R}^3),\Pi\in\mathcal{P}}{\text{find}} \begin{cases} \eta = \Pi\circ\varphi & \text{(reprojection)} \quad (4) \\ \mathcal{T}\varphi = \mathcal{T}\zeta & \text{(deformation)} \quad (5) \end{cases}$$

We observe that the result does not depend on the actual template's embedding but on its metric tensor only.

Proof of proposition 2. The reprojection constraint (4) is obtained by substituting the definition (3) of φ in the reprojection constraint (1). The deformation constraint (5) is obtained by differentiating the definition (3) of φ, giving $\nabla\varphi = (\nabla\Psi\circ\zeta)\nabla\zeta$, and multiplying it by its transpose to the left, eliminating Ψ using the deformation constraint (2).

3.3. Embedding-based Point-Normal Formulation

We now derive a formulation using $\mu(\nabla\varphi)$ (a vector colinear with the surface normal) but not the full tangent plane $\nabla\varphi$. We define $\mu : \mathbb{R}^{3\times 2} \to \mathbb{R}^3$ as $\mu((\mathbf{u}\ \ \mathbf{v})) \overset{\text{def}}{=} \mathbf{u}\times\mathbf{v}$.

Proposition 3 (Point-Normal Formulation of C-IsoSfT)

$$\underset{\varphi\in C^2(\Omega,\mathbb{R}^3),\Pi\in\mathcal{P}}{\text{find}} \begin{cases} \eta = \Pi\circ\varphi & \text{(reprojection)} \quad (6) \\ F[\varphi,\Pi] = \mathbf{0} & \text{(deformation)} \quad (7) \end{cases}$$

with:

$$\begin{aligned} F[\varphi,\Pi] \overset{\text{def}}{=}\ & \nabla\eta\,\mathrm{adj}\,(\mathcal{T}\zeta)\nabla\eta^\top \\ & +(\nabla\Pi\circ\varphi)\mu(\nabla\varphi)\mu(\nabla\varphi)^\top(\nabla\Pi\circ\varphi)^\top \\ & -\det(\mathcal{T}\zeta)(\nabla\Pi\circ\varphi)(\nabla\Pi\circ\varphi)^\top, \end{aligned}$$

where adj *is the adjugate matrix (the transpose of the cofactor matrix,* $\mathrm{adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{A}^{-1}$*).*

Proof of proposition 3. The reprojection constraints (4) and (6) are just the same. As for the deformation constraint (7), we invoke Cholesky decomposition of the metric tensor $\mathcal{T}\zeta$. We define $\Gamma \in C^0(\Omega,\mathbb{R}^{2\times 2})$ as $\Gamma^\top\Gamma \overset{\text{def}}{=} \mathcal{T}\zeta$. The deformation constraint (5) is thus transformed to:

$$\left(\mathcal{T}\varphi = \mathcal{T}\zeta = \Gamma^\top\Gamma\right) \ \Rightarrow\ \left(\Gamma^{-\top}\nabla\varphi^\top\nabla\varphi\,\Gamma^{-1} = \mathbf{I}\right).$$

We differentiate the reprojection constraint (6) and multiply it to the right by Γ^{-1}:

$$\nabla\eta\,\Gamma^{-1} \;=\; (\nabla\Pi\circ\varphi)\nabla\varphi\,\Gamma^{-1}.$$

It can be easily shown that matrix $(\nabla\varphi\,\Gamma^{-1}\;\; \mu(\nabla\varphi\,\Gamma^{-1})) \in C^0(\Omega,\mathbb{O})$. We thus append $\mu(\nabla\varphi\,\Gamma^{-1}))$ (which is the surface normal) as a third column to the equation above:

$$\begin{aligned}&(\nabla\eta\,\Gamma^{-1}\;\; (\nabla\Pi\circ\varphi)\mu(\nabla\varphi\,\Gamma^{-1}))\\ =\;&(\nabla\Pi\circ\varphi)(\nabla\varphi\,\Gamma^{-1}\;\; \mu(\nabla\varphi\,\Gamma^{-1})).\end{aligned}$$

We then multiply each side of the equation by its transpose to the right, yielding:

$$\begin{aligned}(\nabla\Pi\circ\varphi)(\nabla\Pi\circ\varphi)^\top \;&=\; \nabla\eta\,\Gamma^{-1}\Gamma^{-\top}\nabla\eta^\top\\ +(\nabla\Pi\circ\varphi)\mu(\nabla\varphi\,\Gamma^{-1})\mu(\nabla\varphi\,\Gamma^{-1})^\top&(\nabla\Pi\circ\varphi)^\top.\end{aligned}$$

Using $\mu(\nabla\varphi\,\Gamma^{-1}) \;=\; \frac{1}{\det(\Gamma)}\mu(\nabla\varphi)$, multiplying by $\det(\mathcal{T}\zeta)$ and simplifying, we arrive at equation (7). This is the *general equation of C-IsoSfT with general template parameterization*.

4. Local Uncalibrated Solution

We now derive a practical solution to uncalibrated reconstruction which can be computed locally. We use a pinhole camera with unknown focal length $f \in \mathbb{R}^+$. We assume that the other intrinsics (such as the principal point) were 'undone'. We write the projection function as Π_f to emphasize its dependency on f.

4.1. Piecewise Weak-Perspective

Full-Perspective (FP) projection is written $\Pi_f\circ\varphi = \frac{f}{\varphi_Z}\mathbf{S}\varphi$, where φ_Z is the depth, given by the third component of φ and $\mathbf{S} \stackrel{\text{def}}{=} (\mathbf{I}\;\mathbf{0}) \in \mathbb{R}^{2\times 3}$. The Jacobian matrix of FP is spatially varying. IsoSfT can be solved in closed-form with FP [2] but not C-IsoSfT. Weak-Perspective (WP) is a zeroth order approximation of FP obtained by replacing the depth φ_Z by some average $d \in \mathbb{R}^+$, giving $\Pi_f\circ\varphi \approx a\mathbf{S}\varphi$, with $a \stackrel{\text{def}}{=} \frac{f}{d}$ the WP scale factor. One cannot solve for f and d individually with WP in C-IsoSfT but only for their ratio a. The Jacobian matrix of WP is constant and is given by $a\mathbf{S}$.

We here propose the Piecewise WP (PWP) model. The idea is to define a local WP model at each point. In other words, PWP reproduces exactly FP projection but approximates its Jacobian matrix:

$$\Pi_f\circ\varphi \stackrel{\text{def}}{=} \alpha\mathbf{S}\varphi,\;\; \alpha \stackrel{\text{def}}{=} \frac{f}{\varphi_Z} \;\text{ and }\; \nabla\Pi_f\circ\varphi \approx \alpha\mathbf{S}\varphi$$

This is in practice a very good approximation of FP's differential properties. Our method first solves for α as an uncalibrated solution to C-IsoSfT, then calibrates f, and finally returns φ. It can be used for IsoSfT by simply skipping the second step.

4.2. Uncalibrated Solution of C-IsoSfT with PWP

We instantiate the point-normal formulation (proposition 3) with our PWP model. The reprojection constraint (6) becomes $\eta = \alpha\mathbf{S}\varphi$. It allows us to solve for $\varphi_X = \frac{\eta_x}{\alpha}$ and $\varphi_Y = \frac{\eta_y}{\alpha}$ while the deformation constraint (7) allows us to solve for α. Defining $\nu \stackrel{\text{def}}{=} \mathbf{S}\mu(\nabla\varphi)$ as the scaled normal's first two elements, this constraint becomes:

$$\nabla\eta\,\mathrm{adj}\,(\mathcal{T}\zeta)\nabla\eta^\top + \alpha^2\nu\nu^\top - \alpha^2\det(\mathcal{T}\zeta)\mathbf{I} \;=\; \mathbf{0}. \quad (8)$$

Theorem 1 *Equation (8) has a unique solution for α and at most two solutions for ν, each of them corresponding to a solution for the normal ξ, given by:*

$$\alpha \;=\; \sqrt{\lambda_1\left((\mathcal{T}\eta)(\mathcal{T}\zeta)^{-1}\right)} \quad (9)$$

$$\nu_\pm \;=\; \pm\epsilon\left(\det(\mathcal{T}\zeta)\mathbf{I} - \frac{1}{\alpha^2}\nabla\eta\,\mathrm{adj}\,(\mathcal{T}\zeta)\nabla\eta^\top\right) \quad (10)$$

$$\xi_\pm \;=\; \frac{1}{\|\mu(\nabla\zeta)\|_2}\begin{pmatrix}\nu_\pm\\ -\sqrt{\|\mu(\nabla\zeta)\|_2^2 - \|\nu_\pm\|_2^2}\end{pmatrix} \quad (11)$$

Both solutions for the normal are front facing and cannot be disambiguated at this stage. However, they collapse to $\xi_\pm \propto (0\;\;0\;\;1)^\top$ for frontoparallel patches.

We will use the normal as a clue to avoid local degeneracies when estimating the focal length.

Lemma 1 *Let $\mathbf{A} \in \mathbb{R}^{2\times 2}$. The eigenvalues of $\mathbf{A} - \lambda_j(\mathbf{A})\,\mathbf{I}$ are given by $\lambda_i(\mathbf{A} - \lambda_j(\mathbf{A})\,\mathbf{I}) = \lambda_i(\mathbf{A}) - \lambda_j(\mathbf{A})$, implying:*

$$\begin{cases}\lambda_1(\mathbf{A} - \lambda_1(\mathbf{A})\,\mathbf{I}) &=& 0\\ \lambda_2(\mathbf{A} - \lambda_1(\mathbf{A})\,\mathbf{I}) &=& \lambda_2(\mathbf{A}) - \lambda_1(\mathbf{A}) \le 0\\ \lambda_1(\mathbf{A} - \lambda_2(\mathbf{A})\,\mathbf{I}) &=& \lambda_1(\mathbf{A}) - \lambda_2(\mathbf{A}) \ge 0\\ \lambda_2(\mathbf{A} - \lambda_2(\mathbf{A})\,\mathbf{I}) &=& 0.\end{cases}$$

Proof of lemma 1. We replace $\mathbf{A}$ by its eigendecomposition in $\mathbf{A} - \lambda_j(\mathbf{A})\mathbf{I}$:

$$\mathbf{A} - \lambda_j(\mathbf{A})\mathbf{I} = \mathbf{P}\,\mathrm{diag}(\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}))\,\mathbf{P}^\top - \lambda_j(\mathbf{A})\mathbf{I}.$$

Because $\mathbf{P}\mathbf{P}^\top = \mathbf{I}$ we factorize this equation as:

$$\mathbf{P}\,\mathrm{diag}(\lambda_1(\mathbf{A}) - \lambda_j(\mathbf{A}), \lambda_2(\mathbf{A}) - \lambda_j(\mathbf{A}))\,\mathbf{P}^\top,$$

from which we easily conclude.

Lemma 2 *The image of $\mu(\nabla\varphi)$ is colinear with the normal. We also have $\mu(\nabla\varphi) = (\nabla\Psi\circ\zeta)\mu(\nabla\zeta)$, implying $\|\mu(\nabla\varphi)\|_2 = \|\mu(\nabla\zeta)\|_2$.*

Proof of Lemma 2. We substitute $\frac{\partial\varphi}{\partial\star} = (\nabla\Psi\circ\zeta)\frac{\partial\zeta}{\partial\star}$ in $\mu(\nabla\varphi) = \frac{\partial\varphi}{\partial x}\times\frac{\partial\varphi}{\partial y}$. We then use the general rule $(\mathbf{R}\mathbf{u})\times(\mathbf{R}\mathbf{v}) = \det(\mathbf{R})\mathbf{R}^{-\top}(\mathbf{u}\times\mathbf{v})$ and the deformation constraint (2) to finalize the derivation.

Proof of theorem 1. A key step in our proof is rewriting equation (8) as:

$$\nu\nu^\top \propto \alpha^2\mathbf{I} - \Theta \quad \text{with} \quad \Theta \stackrel{\text{def}}{=} \nabla\eta(\mathcal{T}\zeta)^{-1}\nabla\eta^\top,$$

where we simply divided by $\det(\mathcal{T}\zeta) > 0$. Because $\nu\nu^\top$ is positive semi-definite or null, its singular values are respectively not lower than zero and zero. This leads to:

$$\lambda_1\left(\Theta - \alpha^2\mathbf{I}\right) = 0 \quad (12)$$
$$\lambda_2\left(\Theta - \alpha^2\mathbf{I}\right) \leq 0. \quad (13)$$

Equation (12) implies $\det\left(\Theta - \alpha^2\mathbf{I}\right) = 0$, which is the characteristic polynomial of Θ. Therefore, $\exists j \in \{1, 2\}$ such that $\alpha^2 = \lambda_j(\Theta)$. Equation (13) then implies $\alpha^2 = \lambda_1(\Theta)$ using lemma 1. Using $\lambda_j(\mathbf{AB}) = \lambda_j(\mathbf{BA})$ we finally arrive at equation (9).

As for the normal's solutions, we rearrange equation (8) in:

$$\nu\nu^\top = \det(\mathcal{T}\zeta)\mathbf{I} - \frac{1}{\alpha^2}\nabla\eta\,\mathrm{adj}\,(\mathcal{T}\zeta)\nabla\eta^\top \propto \Theta - \alpha^2\mathbf{I}.$$

Because $\alpha^2 = \lambda_1(\Theta)$, lemma 1 shows that the right hand side's image are symmetric rank-1 matrices. We thus obtain ν up to sign as the singular vector associated to the non-zero singular value using ϵ. We find the last element ξ_Z of the scaled normal using $\|\nu\|_2^2 + \xi_Z^2 = \|\xi\|_2^2 = \|\mu(\nabla\zeta)\|_2^2$ from lemma 2, and keep only the negative solution to ensure that the recovered normal is front facing.

5. Focal Length Calibration

Our main result in this section is to compute the focal length analytically from the uncalibrated solution α.

5.1. Basic Equations

Starting from the point-tangent formulation (proposition 2), we use the reprojection constraint (4) to establish:

$$\varphi = \frac{1}{\alpha}\begin{pmatrix}\eta\\ f\end{pmatrix} \quad \text{and} \quad \nabla\varphi = -\begin{pmatrix}\eta\\ f\end{pmatrix}\frac{\nabla\alpha}{\alpha^2} + \frac{1}{\alpha}\begin{pmatrix}\nabla\eta\\ \mathbf{0}^\top\end{pmatrix}.$$

We use this to expand the metric tensor of the embedding:

$$\mathcal{T}\varphi = \frac{1}{\alpha^4}\nabla\alpha^\top\left(\eta^\top\eta + f^2\right)\nabla\alpha + \frac{1}{\alpha^2}\mathcal{T}\eta$$
$$-\frac{1}{\alpha^3}\left(\nabla\alpha^\top\eta^\top\nabla\eta + \nabla\eta^\top\eta\nabla\alpha\right).$$

Plugging this equation in the deformation constraint (5) then leads to:

$$f^2\mathcal{T}\alpha = \alpha^4\mathcal{T}\zeta - \|\eta\|_2^2\mathcal{T}\alpha - \alpha^2\mathcal{T}\eta$$
$$-\alpha\left(\nabla\alpha^\top\eta^\top\nabla\eta + \nabla\eta^\top\eta\nabla\alpha\right). \quad (14)$$

The image of $\mathcal{T}\alpha$ is the set of rank-1 matrices. Equation (14) thus carries two constraints but because one was used to estimate α only one is independent. We multiply the equation by $\nabla\alpha$ to the left and $\nabla\alpha^\top$ to the right. Given that $\nabla\alpha\mathcal{T}\alpha\nabla\alpha^\top = \|\nabla\alpha\|_2^4$ we obtain the following analytical solution for f:

$$f^2 = \frac{\alpha^2}{\|\nabla\alpha\|_2^4}\nabla\alpha\left(\alpha^2\mathcal{T}\zeta - \mathcal{T}\eta\right)\nabla\alpha^\top$$
$$-\frac{\alpha}{\|\nabla\alpha\|_2^2}\left(\eta^\top\nabla\eta\nabla\alpha^\top + \nabla\alpha\nabla\eta^\top\eta\right) - \|\eta\|_2^2. \quad (15)$$

5.2. Geometric Interpretation

Criterion (14) is derived from the isometric deformation constraint. It expresses the fact that at every point, the length of an infinitesimal step in any direction is preserved. To be more general, we can prove that it preserves the length of every 2D curve lying on the template shape. For $b \in \mathbb{R}$ and $\gamma \in C^1([0;1], \Omega)$ some 2D curve, we have that:

$$\int_{[0;1]} \|\nabla\varphi \circ \gamma\|_2^b \,\mathrm{d}t = \int_{[0;1]} \|\nabla\zeta \circ \gamma\|_2^b \,\mathrm{d}t.$$

This is easily shown using the definition (3) of φ and the deformation constraint (2): $\|\nabla\varphi \circ \gamma\|_2^b = \|((\nabla\psi \circ \zeta)\nabla\zeta) \circ \gamma\|_2^b = \|\nabla\zeta \circ \gamma\|_2^b$.

5.3. Degeneracies

Geometry. Degenerate cases arise when the focal length cannot be estimated uniquely from the data. Inspecting criterion (14), we can figure out that a point $\mathbf{q} \in \Omega$ is in a degenerate configuration if $\mathcal{T}\alpha(\mathbf{q}) = \mathbf{0}$, equivalent to $\nabla\alpha(\mathbf{q}) = \mathbf{0}$. In other words, a point is degenerate if the local shape around it is fronto-parallel. Consequently, the data is degenerate if $\mathcal{T}\alpha = 0$, that is to say if the shape is flat and fronto-parallel. This was a known degenerate case in plane-based camera calibration [11].

Detection. In practice, we require that the angle between the normal ξ and $\mathbf{z} \stackrel{\text{def}}{=} (0\ 0\ 1)^\top$ be greater than some minimal angle $r \in \mathbb{R}^+$ for a point to stably contribute to focal length estimation. This criterion is equivalent to:

$$\frac{\|\nu_\pm\|_2^2}{\|\mu(\nabla\zeta)\|_2^2} \geq \sin(r)^2, \quad (16)$$

which can be computed in spite of the two-way ambiguity since $\|\nu_+\|_2 = \|\nu_-\|_2$. The equivalence is proved as follows. Because $|\angle(\xi_\pm, \mathbf{z})| \leq \frac{\pi}{2}$, $\angle(\xi_\pm, \mathbf{z}) > r$ is equivalent to $\cos(\angle(\xi_\pm, \mathbf{z})) \leq \cos(r)$. Squaring and using the dot product, this is rewritten as $(\xi_\pm^\top\mathbf{z})^2 \leq \cos(r)^2$. Using the solution (11) for $\xi_\pm$ allows us to rewrite the left hand side as: $\frac{\|\mu(\nabla\zeta)\|_2^2 - \|\nu\|_2^2}{\|\mu(\nabla\zeta)\|_2^2} = 1 - \frac{\|\nu\|_2^2}{\|\mu(\nabla\zeta)\|_2^2}$ and to arrive at criterion (16).

6. Robust Implementation

We implemented a wide-baseline version of our method using keypoints. Some preprocessing is done off-line on the template, including acquiring its shape if necessary and detecting SIFT keypoints [5]. We standardize the image coordinates to $[0;1]^2$ in the parameterization space and in the input image for numerical stability. Given an input image, the following steps are then taken at runtime:

1. Putative matching. We detect SIFT keypoints and match them with local consistency [9]. This results in a set $\{\mathbf{q}_k \leftrightarrow \mathbf{p}_k\}$ of putative matches with $k = 1, \ldots, N_{\text{putatives}}$. We compute the template's metric tensor for every matched keypoint as $\mathbf{L}_k \stackrel{\text{def}}{=} (\mathcal{T}\zeta)(\mathbf{q}_k)$ and $w_k \stackrel{\text{def}}{=} \|\mu(\nabla\zeta)(\mathbf{q}_k)\|_2^2$.

2. Multi-scale local warps. We use every match to define N_{scales} local warps (typically $N_{\text{scales}} = 10$). For that, we define a set of local scales $\{s_h\}$ evenly from 5% to 50% of the template size with $h = 1, \ldots, N_{\text{scales}}$. At local scale s_h, the support region $\bar{\Omega}_{k,h} \subset \Omega$ is circular with diameter s and centred on the template keypoint $\mathbf{q}_k$. The local scale trades-off stability and deformation complexity: larger scale improves stability but increase sensitivity to high-frequency surface deformation. The local warp $\bar{\eta}_{k,h} : C^2(\bar{\Omega}_{k,h}, \mathbb{R}^2)$ is estimated from all point matches lying in $\bar{\Omega}_{k,h}$. Following [9], we use a fixed number of 9 control centres and a Thin-Plate Spline (TPS) to compute $\mathbf{J}_{k,h} \stackrel{\text{def}}{=} (\nabla\bar{\eta}_{k,h})(\mathbf{q}_k) \in \mathbb{R}^{2\times 2}$ and $\mathbf{H}_{k,h} \stackrel{\text{def}}{=} \mathbf{J}_{k,h}^\top \mathbf{J}_{k,h} = (\mathcal{T}\bar{\eta}_{k,h})(\mathbf{q}_k) \in \mathbb{R}^{2\times 2}$ analytically. We end up with a pool of warps whose size is of the order of $N_{\text{putatives}} N_{\text{scales}}$.

3. Uncalibrated shape and focal length. We estimate the PWP scale factor $a_{k,h}$ and a focal length estimate $f_{k,h}$ for every warp in the pool. The former is obtained through equation (9) as:

$$a_{k,h} \;=\; \sqrt{\lambda_1\left(\mathbf{H}_{k,h}\mathbf{L}_k^{-1}\right)}.$$

The latter is computed through equation (14), which requires an estimate of $\mathbf{d}_{k,h}^\top \stackrel{\text{def}}{=} (\nabla\alpha_{k,h})(\mathbf{q}_k) \in \mathbb{R}^{1\times 2}$. This is obtained by fitting a TPS to an estimate of $\alpha_{k,h}$ at each keypoint in $\bar{\Omega}_{k,h}$ computed using the equation directly above. We arrive at:

$$f_{k,h}^2 \;=\; \frac{a_{k,h}^2}{\|\mathbf{d}_{k,h}\|_2^4}\mathbf{d}_{k,h}^\top\left(a_{k,h}^2\mathbf{L}_k - \mathbf{H}_{k,h}\right)\mathbf{d}_{k,h}$$
$$-\frac{a_{k,h}}{\|\mathbf{d}_{k,h}\|_2^2}\left(\mathbf{p}_k^\top\mathbf{J}_{k,h}\mathbf{d}_{k,h} + \mathbf{d}_{k,h}^\top\mathbf{J}_{k,h}^\top\mathbf{p}_k\right) - \|\mathbf{p}_k\|_2^2.$$

Noise makes the right-hand side negative for a few samples; they are discarded.

4. Robust focal length. We have a large number of candidate focal length estimates. Our robust estimation strategy is to select the f compatible with as many estimates as possible by solving:

$$\hat{f} \;=\; \arg\max_{f\in\mathbb{R}^+} \sum_{k=1}^{N_{\text{putatives}}} \sum_{h=1}^{N_{\text{scales}}} \delta_{k,h}\rho_g(f, f_{k,h})$$

where ρ_g is a step function of width g:

$$\rho_g(f, f_{k,h}) = 1 \text{ if } |f - f_{k,h}| < g \text{ and } 0 \text{ otherwise.}$$

We typically use 1% of the image size for g, and solve the problem by sampling focal length estimates. The indicator $\delta_{k,h}$ test every estimate $f_{k,h}$ against local degeneracy by implementing the test (16) with $r = 5°$:

$$\delta_{k,h} = 1 \text{ if } \|\mathbf{v}_{k,h}\|_2^2 \geq \sin(r)^2 w_k^2 \text{ and } 0 \text{ otherwise,}$$

with $\mathbf{v}_{k,h}$ given by equation (10) as:

$$\mathbf{v}_{k,h} \;\stackrel{\text{def}}{=}\; \epsilon\left(\det(\mathbf{L}_k)\mathbf{I} - \frac{1}{a_{k,h}}\nabla\bar{\eta}_{k,h}\,\text{adj}(\mathbf{L}_k)\bar{\eta}_{k,h}^\top\right).$$

5. Local scale selection and erroneous match pruning. For every putative feature match k, we select the largest local scale s_h whose local focal length estimate $f_{k,h}$ agrees with the robust estimate $\hat{f}$ by testing $\rho_g(\hat{f}, f_{k,h})$. If no scales pass the test, the match is discarded.

7. Experimental Results

7.1. Compared Methods and Measured Errors

We compared 7 methods built on 3 base methods from the 3 categories outlined in §2: *(C1)* PWP (our proposed analytical framework), *(C2)* SLZ (a convex method [10]) and *(C3)* REF (iterative nonlinear refinement [3]). For a method, the leading letter may be U or C: U means that the focal length is estimated by our method or refined and C means that the true focal length (for simulated data) or the focal obtained by static calibration (for real data) is used. For instance, U-PWP is the proposed method from §6, C-SLZ is [10] and C-REF-SHAPE is the nonlinear method in [3]. We measured the average depth error in mm and the relative focal length error in %.

7.2. Simulated Data

We simulated 800×800 squared pixels images of several deformable isometric surfaces [7] with no degeneracies. The default parameters were a focal length of 800 pixels, an image noise of 1.5 pixels and 200 point matches. Each of these 3 parameters was varied on turn and errors were measured over 50 runs for each configuration. The results are shown in figure 2.

We observe in the first column of graphs that the focal length error is always below 10% for the proposed analytical solution U-PWP. It increases with the noise, but decreases for larger numbers of point matches. It has a minimum for a focal length of around 600 pixels. This may be

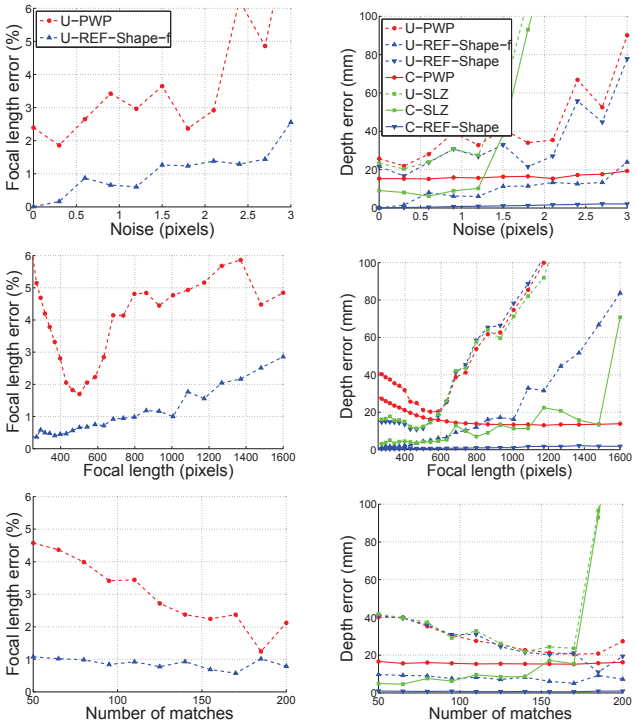


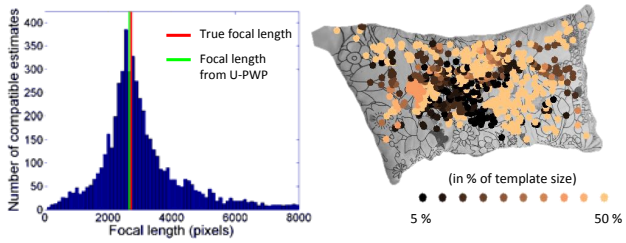
Figure 2. **Results on simulated data.**

explained by the fact that for short focal lengths the PWP approximation tends to be less accurate, while large focal lengths tend to be ill-constrained since they cancel perspective. Shape and focal length refinement by U-REF-SHAPE-F always improve on the results of U-PWP. This is because U-REF-SHAPE-F uses the full-perspective model. It thus does not suffer from the same approximation error. Moreover, it minimizes the reprojection error, which is physically meaningful.

We observe that the depth error of uncalibrated methods increases with the focal length while the error of calibrated methods is approximately steady or decreases. The noise and number of points influence the depth error as expected. All methods are sensitive to the focal length accuracy. We observe that there are three groups of methods. The first one is U-PWP, U-SLZ and U-REF-SHAPE, which use the focal length estimated by U-PWP. They are outperformed by U-REF-SHAPE-F which refines this focal length estimate, and forms the second group. As expected, the third group, made of calibrated methods C-PWP, C-SALZ and C-REF-SHAPE performs better than the two others.

7.3. Real Data

Cushion. This dataset shows a cushion in two different poses. One is used to build the template and the other one to test the methods. The deformation magnitude is 52.27 mm. This dataset was used in figure 1, and more results are shown in figures 3 and 4. We detected 2923 SIFT keypoints in the template and 9472 in the input image, from which we obtained 2923 putative matches, filtered down to 617 after spatial consistency was enforced. For the input image, the groundtruth focal length was 2727.1 pixels. The histogram of local focal length estimates is shown in figure 3. The focal length we estimated with U-PWP is 2668.1 pixels and it is 2801.5 pixels with U-REF-SHAPE-F. This means a relative error of 2.16% and 2.71% respectively. Once the focal length was robustly estimated we took the number of matches down to 612 by checking that their focal length estimate was correct at one scale at least. The selected scale for each match is shown in figure 3. We observe that the isolated matches which were kept have a large local scale, while matches in dense keypoint areas may have a smaller scale, especially if the deformation is important. The reconstructed shape, as well as the groundtruth shape and a color-coded comparison can be seen in figure 4. The average shape errors were (in mm) U-PWP: 16.94, U-SALZ: 15.45, C-PWP: 10.60, C-SALZ: 10.20, U-REF-SHAPE: 8.63, C-REF-SHAPE and U-REF-SHAPE-F: 4.04.



Histogram of local focal length estimates Selected local feature scales

Figure 3. **Results on the Cushion dataset.**

Paper. We tested the 7 methods on the CVLab's paper dataset. This shows a piece of paper being gently bent. The groundtruth shape and constant focal length were provided.

We observe that for U-PWP the focal length error is generally below 10%, while for U-REF-SHAPE-F it is generally below 5%. However, the error may be large at some frames for both methods. These large errors are due to the shape being approximately flat and fronto-parallel at these frames, a degenerate configuration that we identified in §5.3. The depth error is much lower for the calibrated methods than for the uncalibrated methods. All three calibrated methods have the same order of errors. U-REF-SHAPE-F is the best performing of the uncalibrated methods, reaching almost the same accuracy as the calibrated methods, except at those frames where the pose is degenerate.

Figure 4. **Results on the Cushion dataset.** The groundtruth was obtained using dense Rigid Structure-from-Motion.

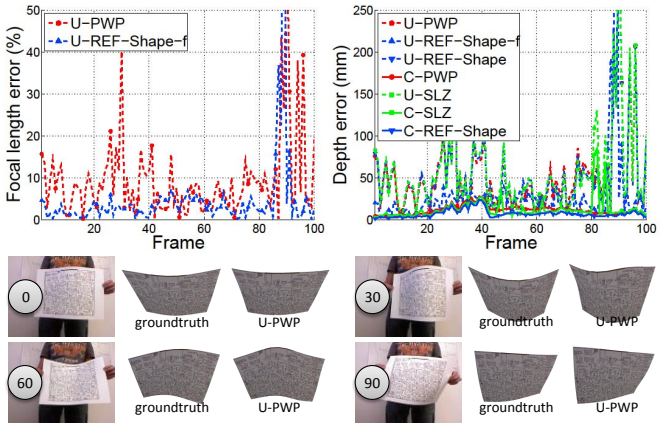


Figure 5. **Results on the CVLab's paper sequence.**

8. Conclusion

We have proposed the first method which solves the Isometric Shape-from-Template problem analytically while recovering the camera's focal length. The proposed method is based on Piecewise Weak-Perspective (PWP), a projection model which approximates perspective projection's partial derivatives with an infinitesimal weak-perspective model. Our experimental results show that the method gives sensible estimates: the focal length error was less than 10% in most cases. We showed that using the true focal length with our PWP model leads to a depth error comparable to state of the art algorithms, including nonlinear refinement of the reprojection error (we recall that the proposed method does not use numerical optimization). The focal length estimate is sensitive to noise in near degenerate configurations. These configurations must be detected to ensure stability, by means of testing a global plane homography, for instance. Our current implementation has not been designed for real-time performance. However, we believe that, being local, our method can run extremely fast on parallel architectures such as GPUs, and thus provide 3D shape in real-time for environments in which the user may need to change the camera's zoom while filming.

Acknowledgements. We thank the authors of [6] for their datasets and the authors of [3, 9, 10] for their code. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE.

References

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

[2] A. Bartoli, Y. Gérard, 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. CVPR, 2012. 1, 2, 3, 4

[3] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres. Monocular template-based reconstruction of smooth and inextensible surfaces. ACCV, 2010. 1, 2, 6, 8

[4] N. A. Gumerov, A. Zandifar, R. Duraiswami, and L. S. Davis. 3D structure recovery and unwarping surfaces applicable to planes. *International Journal of Computer Vision*, 66(3):261–281, 2006. 1, 2

[5] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 6

[6] J. Ostlund, A. Varol, D. Ngo, and P. Fua. Laplacian meshes for monocular 3D shape recovery. ECCV, 2012. 1, 2, 8

[7] M. Perriollat and A. Bartoli. A computational model of bounded developable surfaces with application to image-based 3D reconstruction. *Computer Animation and Virtual Worlds*, 2012. 6

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

[9] D. Pizarro and A. Bartoli. Feature-based non-rigid surface detection with self-occlusion reasoning. *International Journal of Computer Vision*, 97(1):54–70, March 2012. 2, 6, 8

[10] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 33(5), May 2011. 1, 2, 6, 8

[11] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. CVPR, 1999. 5