Template-Based Isometric Deformable 3D Reconstruction with Sampling-Based Focal Length Self-Calibration

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

Abstract

It has been shown that a surface deforming isometrically can be reconstructed from a single image and a template 3D shape. Methods from the literature solve this problem efficiently. However, they all assume that the camera model is calibrated, which drastically limits their applicability.

We propose (i) a general variational framework that applies to (calibrated and uncalibrated) general camera models and (ii) self-calibrating 3D reconstruction algorithms for the weak-perspective and full-perspective camera models. In the former case, our algorithm returns the normal field and camera's scale factor. In the latter case, our algorithm returns the normal field, depth and camera's focal length. Our algorithms are the first to achieve deformable 3D reconstruction including camera self-calibration. They apply to much more general setups than existing methods.

Experimental results on simulated and real data show that our algorithms give results with the same level of accuracy as existing methods (which use the true focal length) on perspective images, and correctly find the normal field on affine images for which the existing methods fail.

1. Introduction

The problem of 3D reconstruction of a deformable surface from monocular video data has been well studied over the past decade. In the *template-based* setup in particular, where a reference 3D view of the surface is known, 3D reconstruction is carried out from 3D to 2D correspondences established between the template and an input image of the surface being deformed. Effective algorithms now exist for the two key steps of image matching [11, 12] and 3D shape inference [1, 2, 3, 5, 7, 8, 10, 14, 15, 16, 17]. Because the reprojection constraints are not sufficient to achieve a single solution, most work use deformation constraints such as isometry [1, 2, 3, 5, 10, 14, 15, 16] and conformity [1, 7], and various other priors such as a learnt shape space [17], multiple local surface patches [15] and other visual cues such a shading [8].

A common requirement of all methods from state of the art is that the *camera's intrinsic parameters be known*. While this has initially been a reasonable assumption, being able to self-calibrate the camera would grant 3D reconstruction much more flexibility. In rigid Structure-from-Motion, camera self-calibration is well-understood. The most interesting scenario, both in terms of stability and applicability, is where all the intrinsics are known but the focal length which is also allowed to vary in time [13]. This lets the user to zoom in and out while filming.

This paper proposes a comprehensive framework for 3D reconstruction from a single *uncalibrated* image under isometric surface deformation. In this context, most existing methods use a fully calibrated perspective camera model [1, 2, 3, 7, 8, 10, 14, 15, 16, 17] and are defeated by affine imaging conditions. The reason is that they do not fully exploit the differential surface constraints, and use the so-called *maximum depth heuristic* [10], consisting in maximizing the surface's depth while bounding surface extension [2, 3, 10, 14, 15, 16]. Two exceptions are [1, 5] which use a variational framework with a perspective and an orthographic projection model respectively. In contrast, our general variational framework applies to a general camera model, whether calibrated or uncalibrated. It relates the template to input image warp to the unknown surface embedding. It leads to a general PDE for isometric 3D reconstruction with the camera's intrinsics as free parameters. We establish that in the affine case, only the surface normal can be computed but not the absolute depth, while in the perspective case, both the surface normal, absolute depth and focal length can be estimated. We give two algorithms. Our first algorithm is dedicated to the weak-perspective camera. It computes the surface normal and the camera's scale factor (the ratio between the camera's focal length and the surface's average depth). Our second algorithm is dedicated to the full-perspective camera. It computes the surface normal and depth, and the camera's focal length. Both proposed algoritms are extremely fast.

Experimental results support the fact that focal length self-calibration is feasible. A relative error of a few percents can be reached in most camera/surface configurations, leading to satisfying 3D reconstructions.

Paper organization. §2 reviews state of the art. §3 gives our notation, the problem setup and its modeling. §4 derives our general variational framework for isometric 3D reconstruction. §§5 and 6 specialize this framework to weak-perspective and full-perspective projection respectively, and give solution algorithms for 3D reconstruction including camera self-calibration. §7 reports experimental results and §8 gives conclusions.

2. State of the Art

Reconstructing a deforming surface in the template-based setting has two main steps: input image to template registration and 3D shape inference. The registration step has been effectively solved using feature-based [11, 12] and pixel-based approaches [12]. This paper specifically focuses on the 3D shape inference step under isometric surface deformation [1, 2, 3, 5, 10, 14, 15, 16]. Most of these methods use a convex relaxation of the original problem. The most successful relaxation [2, 14] has been the maximum depth heuristic [10] that consists in maximizing the surface's depth under inextensibility constraints [2, 14] using Second-Order Cone Programming (SOCP). The fastest results were however obtained by solving a variational formulation exploiting the differential structure of local isometry in the perspective [1] and orthographic [5] projection cases.

All the previously cited methods make a fundamental assumptions: the camera model is perspective projection and its intrinsics are known (except [5] which uses orthographic projection). These methods are defeated by affine imaging conditions since they do not directly exploit the problem's full differential structure. In other words, they only compute depth, which is not recoverable in affine imaging conditions.

We generalize the previous variational formulations [1, 5] to an arbitrary projection function. We specifically instanciate our formulation for weak-perpective and full-perspective projection. In the former case, our algorithm computes the scale factor and the surface normal. In the latter case, our algorithm computes the camera's focal length, the surface normal and depth. Our method is the first to solve 3D deformable shape reconstruction while performing camera self-calibration.

3. Notation and Modeling

Our notation and modeling are illustrated in figure 1. The *template domain* is written $\Omega \subset \mathbb{R}^2$. The unknown 3D surface is parameterized by an isometric embedding of the template, represented by the *surface embedding function* $\varphi : \Omega \to \mathbb{R}^3$. The *camera projection function* is written $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$. The known *template-to-image warp function* is written $\eta : \Omega \to \mathbb{R}^2$. Finally, the unknown *surface unit normal function* is written $\xi : \Omega \to \mathbb{S}^3$. We use the notation $\mathsf{J}_f \stackrel{\text{def}}{=} \frac{\partial f}{\partial \mathbf{p}}$ for the Jacobian-matrix function of function $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ with $\mathsf{J}_f : \mathbb{R}^d \to \mathbb{R}^{d' \times d}$.

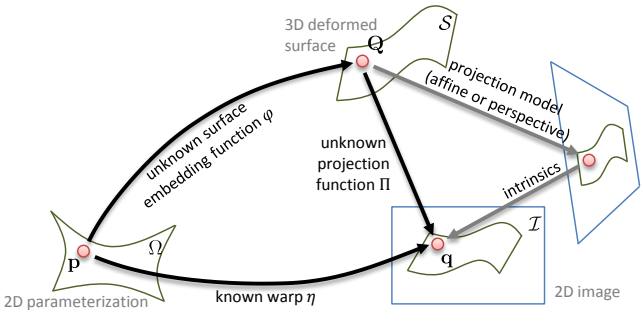

Figure 1. **Modeling monocular template-based surface reconstruction.** The warp η is estimated at the image registration step. Existing reconstruction methods compute the surface embedding function φ assuming that the camera projection function Π is known. In our framework, we estimate both φ and Π. This implies estimating the weak-perspective camera's scale or self-calibrating the full-perspective camera's focal length.

We have three basic constraints. First, composing the surface embedding and camera projection gives the warp; this is the *reprojection constraint*:

$$\eta \;=\; \Pi \circ \varphi. \tag{1}$$

Second, the surface embedding's Jacobian matrix function $\mathsf{J}_\varphi : \Omega \to \mathbb{R}^{3\times 2}$ must be scaled orthonormal for the surface deformation to be isometric; this is the *deformation constraint*:

$$\frac{1}{\lambda}\mathsf{J}_\varphi \;\in\; \tilde{\mathbb{O}}, \tag{2}$$

where $\lambda : \Omega \to \mathbb{R}$, $\lambda > 0$ is the known local scaling function[1] and $\tilde{\mathbb{O}}$ is the set of (3×2) column-orthonormal matrices. Third, the first order partial derivatives of the reprojection constraint must agree; this is the *differential constraint*, generalizing the Sub-Stiefel matrix constraint [4]:

$$\mathsf{J}_\eta \;=\; (\mathsf{J}_\Pi \circ \varphi)\, \mathsf{J}_\varphi, \tag{3}$$

where $\mathsf{J}_\eta : \Omega \to \mathbb{R}^{2\times 2}$ and $\mathsf{J}_\Pi : \mathbb{R}^3 \to \mathbb{R}^{2\times 3}$ are the warp and camera projection's Jacobian-matrix functions respectively.

4. General Isometric 3D Reconstruction

We start from the differential constraint (3), and append the scaled unit surface normal $\lambda\xi$ as the rightmost column of this matrix equality:

$$\begin{pmatrix}\mathsf{J}_\eta & \lambda(\mathsf{J}_\Pi \circ \varphi)\xi\end{pmatrix} \;=\; (\mathsf{J}_\Pi \circ \varphi)\begin{pmatrix}\mathsf{J}_\varphi & \lambda\xi\end{pmatrix}.$$

[1] For a developable surface $\lambda = 1$ while otherwise λ may be the local scaling due to the 2D parameterization Ω obtained by flattening of a 3D template [1].

We multiply each side of this equation to the right by its transpose. Because the deformation constraint (2) implies $\begin{pmatrix}\mathsf{J}_\varphi & \xi\end{pmatrix}\begin{pmatrix}\mathsf{J}_\varphi & \xi\end{pmatrix}^\top = \lambda^2\mathrm{I}$, with I the identity matrix (here of size (3×3)), the equation is simplified, and gives the *general equation of isometric 3D reconstruction*:

$$\mathsf{J}_\eta\mathsf{J}_\eta^\top + \lambda^2(\mathsf{J}_\Pi\circ\varphi)\xi\xi^\top(\mathsf{J}_\Pi\circ\varphi)^\top \;=\; \lambda^2(\mathsf{J}_\Pi\circ\varphi)(\mathsf{J}_\Pi\circ\varphi)^\top. \tag{4}$$

This is a nonlinear PDE in the camera projection Π, the surface embedding φ and its normal ξ. It is clear that ξ may be derived from φ; however, relaxing this dependency leads to a local and computationally fast solution [1]. Due to symmetry, there are only three distinct equations out of the four equations of this matrix equality. This PDE must be solved jointly with the reprojection constraint (1). Because of its global and parametric nature, camera projection will turn into a set of free unknown parameters when specializing this PDE to a particular camera model.

5. Weak-Perspective Solution

We show how the general reconstruction equation (4) is specialized and solved for weak-perspective projection.

5.1. Specializing the General Equation

The general affine camera's projection function is $\Pi_A(\mathbf{Q}) = \mathsf{K}_A\mathsf{S}_A\mathbf{Q}$. In this equation $\mathsf{S}_A = (\mathrm{I}\ \mathbf{0}) \in \mathbb{R}^{2\times 3}$ is a constant matrix and $\mathsf{K}_A \in \mathbb{R}^{2\times 2}$ is an upper triangular matrix containing the camera's three intrinsics. We thus obtain $\mathsf{J}_{\Pi_A} = \mathsf{K}_A\mathsf{S}_A$, that we substitute in the general reconstruction equation (4) to get:

$$\mathsf{J}_\eta\mathsf{J}_\eta^\top + \lambda^2\mathsf{K}_A\mathsf{S}_A\xi\xi^\top\mathsf{S}_A^\top\mathsf{K}_A^\top \;=\; \lambda^2\mathsf{K}_A\mathsf{S}_A\mathsf{S}_A^\top\mathsf{K}_A^\top.$$

Function φ disappears since the affine camera's Jacobian is constant. Defining $\bar\xi : \Omega \to \mathbb{R}^2$, the function giving the first two elements of the unit normal, as $\bar\xi = \mathsf{S}_A\xi$, we get the *affine equation of isometric 3D reconstruction*:

$$\mathsf{J}_\eta\mathsf{J}_\eta^\top + \lambda^2\mathsf{K}_A\bar\xi\bar\xi^\top\mathsf{K}_A^\top = \lambda^2\mathsf{K}_A\mathsf{K}_A^\top.$$

For a weak-perspective camera, $\mathsf{K}_A = \alpha\mathrm{I}$ where the unknown scale $\alpha \stackrel{\text{def}}{=} \frac{f}{d} > 0$ is the ratio between the camera's focal length f and the surface's average depth d. This leads to the *weak-perspective equation of isometric 3D reconstruction*:

$$\mathsf{J}_\eta\mathsf{J}_\eta^\top + \lambda^2\alpha^2\bar\xi\bar\xi^\top = \lambda^2\alpha^2\mathrm{I}. \tag{5}$$

This is a polynomial first-order PDE with $\alpha \in \mathbb{R}_+$ as free parameter.

5.2. Solving

Equation (5) involves function $\bar\xi$ and parameter α. The latter is a free parameter involved globally over the domain Ω by the PDE. Attempting to solve for $\bar\xi$ and α simultaneously leads to a large and untractable polynomial optimization problem. We propose a solution that first computes α globally and then $\bar\xi$ locally.

Solving for α. We rearrange equation (5) as $\lambda^2\alpha^2\bar\xi\bar\xi^\top = \lambda^2\alpha^2\mathrm{I} - \mathsf{J}_\eta\mathsf{J}_\eta^\top$. Because the left-hand side of the equation is a rank-1 matrix, we can write that the right-hand side's determinant vanishes, giving, with $\mu \stackrel{\text{def}}{=} \alpha^2$:

$$\det\left(\lambda^2\mu\mathrm{I} - \mathsf{J}_\eta\mathsf{J}_\eta^\top\right) \;=\; 0. \tag{6}$$

This shows that two solutions for μ could be easily found from the eigenvalues of $\mathsf{J}_\eta\mathsf{J}_\eta^\top$. However, we do not want to solve for μ locally but globally, taking measurements into account over the whole domain Ω, for stability purposes. Expanding equation (6), we get the following degree-two polynomial in μ:

$$\lambda^4\mu^2 - \lambda^2 t\mu + g \;=\; 0, \tag{7}$$

with $t \stackrel{\text{def}}{=} \operatorname{tr}\left(\mathsf{J}_\eta\mathsf{J}_\eta^\top\right)$ and $g \stackrel{\text{def}}{=} \det\left(\mathsf{J}_\eta\mathsf{J}_\eta^\top\right)$. We define the optimization problem to get an estimate of μ from all measurements as:

$$\min_{\mu\in\mathbb{R}} \int_\Omega \left(\lambda^4\mu^2 - \lambda^2 t\mu + g\right)^2 \mathrm{d}\mathbf{p}. \tag{8}$$

Nullifying the cost's μ-derivative yields the following degree-three polynomial:

$$\begin{aligned} 2\mu^3\int_\Omega \lambda^8\,\mathrm{d}\mathbf{p} - 3\mu^2\int_\Omega \lambda^6 t\,\mathrm{d}\mathbf{p} \\ +\mu\int_\Omega \lambda^4\left(2g + t^2\right)\mathrm{d}\mathbf{p} - \int_\Omega \lambda^2 tg\,\mathrm{d}\mathbf{p} \;=\; 0. \end{aligned} \tag{9}$$

We finally solve for μ by keeping the real positive root minimizing the cost function[2] and set $\alpha = \sqrt{\mu}$.

Solving for ξ. We rewrite equation (5) as $\lambda^2\bar\xi\bar\xi^\top = \mathsf{M}$ where $\mathsf{M} \stackrel{\text{def}}{=} \lambda^2\mathrm{I} - \frac{1}{\mu}\mathsf{J}_\eta\mathsf{J}_\eta^\top$. We simply use a rank-one decomposition $\zeta\zeta^\top$ of M, where $\zeta \in \mathbb{R}^2$, that we compute from a Singular Value Decomposition $\mathsf{M} = \mathsf{U}\Sigma\mathsf{U}^\top$ as $\zeta = \frac{1}{\sqrt{\sigma}}\mathbf{u}$, where $\mathbf{u}$ is the column of U associated to the largest singular value σ. We therefore get two solutions $\bar\xi = \pm\frac{\zeta}{\lambda}$. We finally use the constraint $\|\xi\|_2 = 1$, leading to $\xi_Z = -\sqrt{1 - \frac{1}{\lambda^2\sigma}}$ to get two solutions:

$$\xi_1 \;=\; \begin{pmatrix}\zeta \\ -\sqrt{1-\frac{1}{\lambda^2\sigma}}\end{pmatrix} \qquad \text{and} \qquad \xi_2 \;=\; -\begin{pmatrix}\zeta \\ \sqrt{1-\frac{1}{\lambda^2\sigma}}\end{pmatrix}, \tag{10}$$

[2] At least one root of equation (9) is real positive. The quadratic (7) opens upwards (because $\lambda^4 > 0$) and its two roots are real positive (because $t > 0$ and $g \geq 0$). Therefore, the integrand (8) is a positive quartic that has two pairs of repeated real positive roots and is strictly decreasing for $\mu < 0$. The integral cost (8) has thus at least one real positive root.

where, because $\xi_Z < 0$, both possible normal are directed towards the camera. Criteria such as surface integrability or smoothness [3] can be used (through normal integration) to recover a C^1 shape up to scale while disambiguating the normal field. This may however leave some convex/concave ambiguities unresolved.

6. Full-Perspective Solution

We here show how the general reconstruction equation (4) is specialized and solved for full-perspective projection.

6.1. Specializing the General Equation

The general perspective camera's projection function is $\Pi_P(\mathbf{Q}) = \frac{1}{Q_Z}\mathsf{K}_P\mathbf{Q}$. In this equation $\mathsf{K}_P \in \mathbb{R}^{2\times 3}$ contains the five camera's intrinsics. We partition it as $\mathsf{K}_P = \begin{pmatrix}\bar{\mathsf{K}}_P & \mathbf{q}_0\end{pmatrix}$ where $\bar{\mathsf{K}}_P \in \mathbb{R}^{2\times 2}$ is an upper triangular matrix containing the focal length f, the skew τ and the aspect ratio ρ, and $\mathbf{q}_0 \in \mathbb{R}^2$ is the principal point. We thus obtain $\mathsf{J}_{\Pi_P} = \bar{\mathsf{K}}_P\mathsf{S}_P$ with $\mathsf{S}_P \stackrel{\text{def}}{=} \frac{1}{Q_Z}\begin{pmatrix}\mathrm{I} & -\frac{1}{Q_Z}\bar{\mathbf{Q}}\end{pmatrix}$ and $\bar{\mathbf{Q}}^\top \stackrel{\text{def}}{=} (Q_X \;\; Q_Y)$. Substituting Π_P into the reprojection constraint (1) we get:

$$\eta \;=\; \frac{1}{\varphi_Z}\mathsf{K}_P\varphi = \frac{1}{\varphi_Z}\bar{\mathsf{K}}_P\bar{\varphi} + \mathbf{q}_0$$

with $\bar{\varphi}^\top \stackrel{\text{def}}{=} (\varphi_X \;\; \varphi_Y)$. This leads to $\bar{\varphi} = \varphi_Z\bar{\mathsf{K}}_P^{-1}(\eta - \mathbf{q}_0)$. Substituting this expression in $\mathsf{J}_{\Pi_P}\circ\varphi$ gives:

$$\mathsf{J}_{\Pi_P}\circ\varphi \;=\; \frac{1}{\varphi_Z}\begin{pmatrix}\bar{\mathsf{K}}_P & \mathbf{q}_0 - \eta\end{pmatrix}.$$

Finally, substituting this expression in the general reconstruction equation (4) we obtain the *full-perspective equation of isometric 3D reconstruction*:

$$\mathsf{J}_\eta\mathsf{J}_\eta^\top + \frac{\lambda^2}{\varphi_Z^2}\left(\bar{\mathsf{K}}_P\bar{\xi}\bar{\xi}^\top\bar{\mathsf{K}}_P^\top + \xi_Z^2\tilde{\eta}\tilde{\eta}^\top\right) = \frac{\lambda^2}{\varphi_Z^2}\left(\bar{\mathsf{K}}_P\bar{\mathsf{K}}_P^\top + \tilde{\eta}\tilde{\eta}^\top\right),$$

with $\tilde{\eta} = \mathbf{q}_0 - \eta$. We further specialize this equation under the assumption that only the focal length f is unknown and the effect of the other intrinsics were undone. This leads to $\bar{\mathsf{K}}_P = f\mathrm{I}$ and $\mathbf{q}_0 = \mathbf{0}$. Setting $\gamma = \varphi_Z^2$, we obtain:

$$\begin{aligned}\gamma\mathsf{J}_\eta\mathsf{J}_\eta^\top + \lambda^2 f^2\bar{\xi}\bar{\xi}^\top + \lambda^2\xi_Z^2\eta\eta^\top + f\xi_Z\left(\bar{\xi}\eta^\top + \eta\bar{\xi}^\top\right)& \\ = \;\lambda^2 f^2\mathrm{I} + \lambda^2\eta\eta^\top.& \quad (11)\end{aligned}$$

This is a nonlinear PDE with $f \in \mathbb{R}$ as free parameter.

6.2. Solving: Finding an Initialization

Equation (11) involves functions ξ and γ, and parameter f. The latter is involved globally. The relationship between ξ and γ is quite complex and we thus cannot directly exploit it to solve the variational equation efficiently. Indeed, γ gives the depth and with the reprojection constraint, determines function φ, whose first partial derivatives lead to the normal function ξ.

We propose the following estimation procedure: *(i)* sample f over a range of admissible values, *(ii)* for each candidate f value, solve the equation of isometric 3D reconstruction (11) and *(iii)* keep the value of f which best satisfies the global isometric constraint. This procedure makes the solution of step *(ii)* pointwise, easily parallelizable on the GPU and thus extremely fast. We sample 100 f values on a log-scale from 10^2 pixels to 5×10^3 pixels. Note that the template camera's focal length is generally unrelated to the runtime camera's (for instance with printed paper we use the digital texture image as a template).

Solving for γ and ξ given f. We propose an alternative solution to equation (11) to the existing closed-form [1]. Relaxing the constraint relating γ to ξ, equation (11) can be viewed as a system of three polynomials of degree two in four variables at each point $\mathbf{p} \in \Omega$, to which we add the constraint $\|\xi\|_2^2 = 1$. Because of its special structure, this system has at most four solutions. More specifically, there are four solutions for the normal $\bar{\xi}$ and two for the depth γ. However, it has been shown that with this relaxation only one solution satisfies $\gamma > 0$ [1]. Without loss of generality, we here assume $\lambda = 1$ (the surface is developable), but the method applies to an arbitrary local scale function λ. The four considered polynomial constraints depend on six monomials, $\xi_X^2, \xi_Y^2, \xi_Z^2, \xi_X\xi_Y, \xi_Z$ and γ. It can thus be linearized by introducing the following four variables: $\zeta_1 = \xi_X^2$, $\zeta_2 = \xi_Y^2$, $\zeta_3 = \xi_Z^2$ and $\zeta_4 = \xi_X\xi_Y$. With these, we obtain a linear system with four equations and six variables. Our solution method finds the two-dimensional linear subspace of solutions, and selects the four solutions for ξ and the two solutions for γ from the quadratic constraints $\zeta_1\zeta_2 - \zeta_4^2 = 0$ and $\xi_Z^2 = \zeta_3$. We finally keep the only solution such that $\gamma > 0$. We do not keep the two ambiguous solutions for the normal field but rather recompute it a posteriori from function γ.

Selecting the best f sample. The best f sample is selected using global isometry as a criterion. The latter is measured using the so-called Euclidean approximation to geodesics. Let $(\mathbf{p},\mathbf{p}') \in \mathcal{H} \subset \Omega^2$ be a pair of neighboring template points. For this point pair we measure the amount of surface extension or shrinking with respect to the template as:

$$|\delta(\mathbf{p},\mathbf{p}') - \delta(\varphi(\mathbf{p}),\varphi(\mathbf{p}'))|,$$

where $\delta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the Euclidean distance function. We then robustify the criterion by keeping the median value over $\mathcal{H}$:

$$F[\varphi] \;\stackrel{\text{def}}{=}\; \underset{(\mathbf{p},\mathbf{p}')\in\mathcal{H}}{\text{median}}\; |\delta(\mathbf{p},\mathbf{p}') - \delta(\varphi(\mathbf{p}),\varphi(\mathbf{p}'))|. \quad (12)$$

In practice we use $K = 5$ nearest neighbors to construct $\mathcal{H}$ from the input point correspondences.

6.3. Solving: Direct Nonlinear Refinement

We finally propose a variational formulation of the problem and to solve it numerically:

$$\min_{\varphi, f} \int_\Omega \|\eta - \Pi_P \circ \varphi\|_2^2 \, \mathrm{d}\mathbf{p} + \nu \int_\Omega \left\| \mathrm{J}_\varphi^\top \mathrm{J}_\varphi - \lambda^2 \mathrm{I} \right\|_\mathcal{F}^2 \, \mathrm{d}\mathbf{p},$$

with $\nu \in \mathbb{R}_+$ a weight on the isometry constraint, that we here choose empirically. The surface embedding function φ is represented as a linear interpolant of control points positionned on a regular grid. The problem is finally solved using Gauss-Newton. We implemented two versions of the direct nonlinear refinement. The first one uses the static calibration value for f (and is equivalent to the nonlinear method of [2]) while the second one estimates f numerically. The initial solution is provided by our focal length sampling algorithm.

7. Experimental Results

7.1. Compared Methods and Measured Errors

We compared 8 methods: 5 use ground-truth static calibration and 3 perform self-calibration. We measured three types of error: the *f-error* (the relative absolute difference between the true and the estimated focal lengths, in %); the *depth-error* (the average depth discrepancy between the true and the estimated surfaces, in pixels) and the *normal-error* (the average angle between the true and the estimated normal, in degrees – for the weak-perspective solution we use the normal giving the smallest error). Note that for the weak-perspective solution only the normal error is computable. We implemented the registration step as follows, unless stated otherwise. We first used SIFT [6] to obtain putative keypoint correspondences from which we then estimate a Thin-Plate Spline warp η using a robust method based on spatial consistency [12]. We always use the pixel grid in the template to discretize the PDEs.

Methods using static calibration. It should be noted that these compared methods assume the focal length to be known. In the case of simulated data this is the groundtruth focal length; it the case of real data it is obtained from static calibration. STAT-PE is an iterative method using the maximum depth heuristic [10]. STAT-SA is a convex solution using the maximum depth heuristic [14]. STAT-BR is a convex SOCP solution using the maximum depth heuristic [2]. STAT-BA is an analytical solution using variational calculus [1]. STAT-RE is a nonlinear refinement method [2].

Methods performing self-calibration. We compared three proposed methods. SELF-WP is the proposed weak-perspective method of §5. SELF-FP is the proposed full-perspective method of §6.2. SELF-RE is the proposed nonlinear refinement method of §6.3.

7.2. Simulated Data

We used a paper model [9] to simulate isometrically deforming surfaces. We randomly drew m points on the simulated surfaces and projected them with a perspective camera. For each tested configuration, we averaged the results over 50 trials. We varied the simulated focal length (default: 400 pixels), the number of keypoint correspondences (default: 200) and the standard deviation of the gaussian-distributed correspondence noise (default: 1.5 pixels). Our results are displayed in figure 2.

The top row shows the f-error. We observe that SELF-FP degrades with increasing focal length and correspondence noise and improves with increasing number of correspondences. However, the f-error is kept below about 15% and is of a few percents for most simulated configurations. We observe that SELF-RE is kept to less than 1% error for all configurations. This is comparable with an excellent static camera calibration. This means two things: first that SELF-RE's cost function leads to accurate estimates and second that SELF-FP provides SELF-RE with an initialization that allows it to reach an accurate estimate.

The middle row shows the depth error. As with the f-error, we observe that SELF-FP degrades with increasing focal length and correspondence noise and improves with increasing number of correspondences. We observe that SELF-FP is in the range of error of methods using static calibration. SELF-FP allows SELF-RE to converge to a solution which is almost as accurate as STAT-RE, which, because it uses static calibration, we can consider as a lower bound on the error achievable by self-calibration.

The bottom row shows the normal error. We make the same observations for SELF-FP as for the depth error. The normal error is kept between 10–40 degrees. SELF-WP has errors between 7–13 degrees. It gives normal estimates which are almost always more accurate than SELF-FP's despite the significant amount of perspective in short focal length simulated configurations. The curves for SELF-RE and STAT-RE are indistinguisable and lie at around a few degrees error.

The first column shows the result when changing the simulated focal length. When it gets large the imaging function gets closer to parallel projection. Therefore, the accuracy of some methods based on the maximum depth heuristic (STAT-SA and STAT-PE) degrades significantly. The focal length and depth also become ill-constrained as only their ratio can be measured, explaining why we observe that their estimates by SELF-FP degrades. The surface normal however is still well-constrained, as can be observed from SELF-FP's normal estimates. This can be seen from equation (**??**):

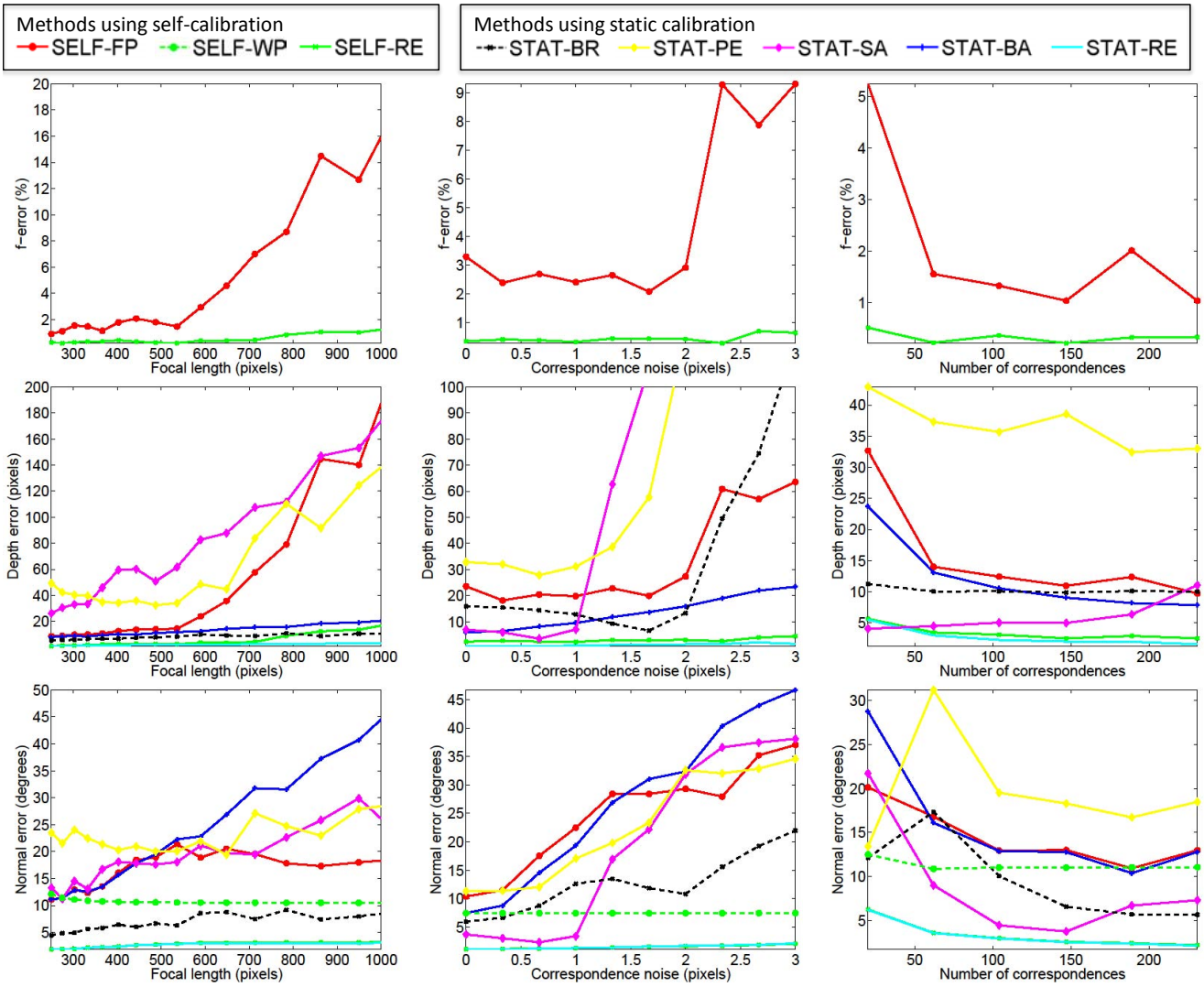

Figure 2. **Experimental results with simulated data.**

when f grows large the depth γ becomes ill-constrained but not the normal ξ. We observe that the f-error and the depth error for SELF-RE increase with the focal length but much less than for SELF-FP, while the normal error is kept to its lower bound provided by STAT-RE.

7.3. Real Data

We tested the above mentioned algorithms on several real datasets. We show results for three selected datasets.

The 9-zoom dataset. This new dataset consists of 9 sets of 3 still images each. Each image shows a deformation of a paper sheet. Each set has a different but constant level of zoom. They thus cover 9 levels of zoom, from small to large. We use two examples from this dataset to illustrate the whole reconstruction pipeline with the key steps shown in figure 3. These two examples respectively use a short and long focal length. We now describe the short focal length example in details. 1432 and 1462 SIFT keypoints [6] were extracted from the template and the input image respectively. 798 putative correspondences were obtained, and 698 were kept after spatial consistency was enforced [12]. A Thin-Plate Spline warp was then fitted to the keypoint correspondences. The 3D reconstructions obtained using static calibration and self-calibration are visually indistinguishable.

The bar-plot in figure 3 shows the true and estimated focal length as the level of zoom varies. From level 1 to level 5 (1831 to 3605 pixels) the f-error is kept below 10% for both SELF-FP and SELF-RE. From level 6 to level 9 (4015 to 5670 pixels) the f-error grows up to 28% for SELF-FP. It however allows SELF-RE to converge to a solution were

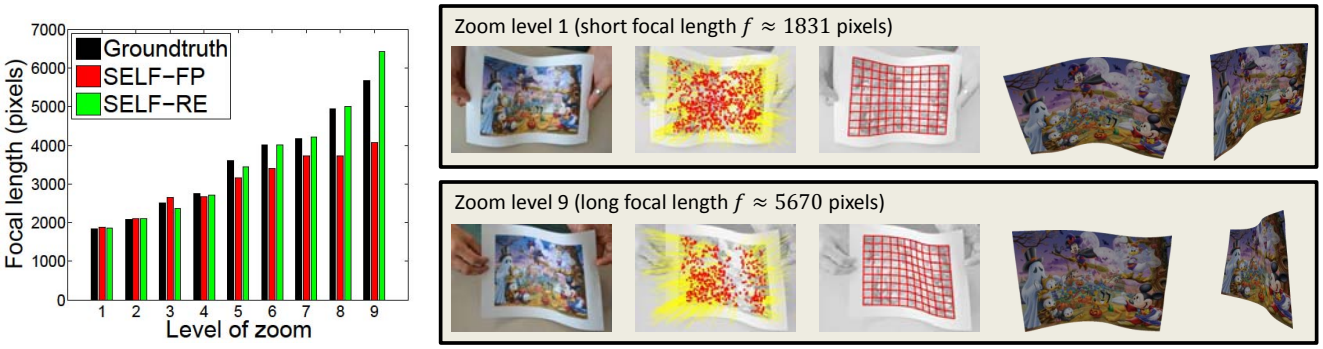

Figure 3. **Results on the 9-zoom dataset.**

the f-error ranges from a few percents (levels 6 to 8) to 13% (level 9). This is a reasonable accuracy given that self-calibration is here performed from noisy keypoint correspondences obtained automatically on a single image.

CVLab's paper sequence dataset. This dataset with estimated groundtruth depth was kindly provided by EPFL's CVLab on their website. We show results for the first 90 frames of this video sequence showing a piece of paper being manually deformed. We used the SIFT correspondences provided with the sequence. The focal length is fixed and its groundtruth value is 528 pixels. Figure 4 shows the results we obtained.

We observe on the left graph that SELF-FP produces a depth error slightly larger that the other methods, but of the same order of magnitude. On the other hand, SELF-RE achieves a depth error comparable to methods using static calibration. The middle graph shows that both SELF-FP and SELF-RE overestimate the focal length by a few dozens of pixels. The right graph shows that the f-error is kept below 12% for SELF-FP and below 8% for SELF-RE. The average f-error is 5.8% and 3.1% for these two methods respectively, which we consider as an accurate result.

The cap dataset. Results for this new dataset are in figure 5. The template here is in 3D since it is non-developable (the cap cannot be isometrically flattened to a plane). The input image shows the cap with a crease in the centre. The groundtruth focal length from static calibration is 2040 pixels. We here followed a special reconstruction procedure in two steps. Because the cap is a 3D object, is it never entirely visible in an input image. Specifically, the textured visible part of the cap is inside the dashed red curve in figure 5. We first reconstructed the visible part of the cap using template-based deformable 3D reconstruction. This was based on 241 semi-automatically established keypoint correspondences. We then transferred the hidden part of the cap from the template by extrapolating the transformation obtained for the reconstructed visible part. We observe that the shape reconstructed by STAT-RE is visually extremely similar to the one reconstructed by SELF-RE. The average relative error to the groundtruth shape obtained by structured lighting is 0.60% and 0.74% for STAT-RE and STAT-FP with standard deviation 0.51% and 0.64% respectively. The estimated focal length was 1890 pixels for SELF-FP and 2118 pixels for SELF-RE, which means an f-error of 7.3% and 3.8% respectively. We consider this as a very successful result.

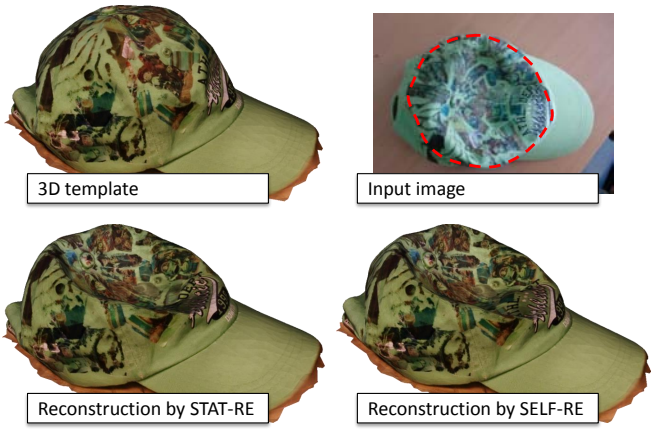

Figure 5. **Results on the cap dataset.**

8. Conclusion

The main conclusion of our paper is that focal length self-calibration in template-based isometric deformable 3D reconstruction is feasible. This is taking the level of flexibility of this type of methods a step further. Our initialization algorithm facilitates accurate 3D reconstruction for small to medium focal length values while our nonlinear refinement algorithm handles small to large focal length values extremely well, being as accurate as methods using static calibration. When the focal length grows too large it cannot

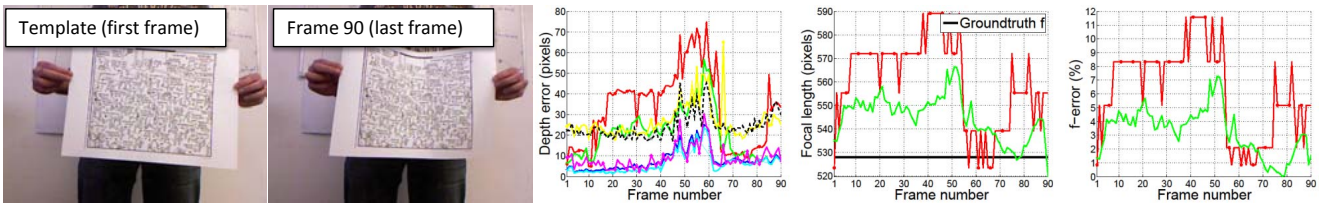

Figure 4. **Results on the paper sequence.** See figure 2 for the graphs' legend.

be computed. We showed how the surface normal can however still be accurately estimated with the weak-perspective projection model. The proposed algorithms were implemented in pure Matlab. They process a few frames per second on a regular PC. Because they are highly parallelizable (except SELF-RE), it is likely that a GPU-C/C++ implementation would process hundreds of frames per second. Future work may address the well-posedness of f computation (a trivial degeneracy for instance is a flat and frontoparallel surface) and the conformal deformation case.

Acknowledgements. This research has received funding from the EU's FP7 through the ERC research grant 307483 FLEXABLE. Thanks to Sebastian Haner for spotting a mistake in the derivation of the local solution in the full-perspective case.

References

[1] 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. In *International Conference on Computer Vision and Pattern Recognition*, 2012. 1, 2, 3, 4, 5

[2] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres. Monocular template-based reconstruction of smooth and inextensible surfaces. In *Asian Conference on Computer Vision*, 2010. 1, 2, 5

[3] A. Ecker, K. Kutulakos, and A. Jepson. Semidefinite programming heuristics for surface reconstruction ambiguities. In *European Conference on Computer Vision*, 2008. 1, 2, 4

[4] R. Ferreira, J. Xavier, and J. Costeira. Shape from motion of nonrigid objects: the case of isometrically deformable flat surfaces. In *British Machine Vision Conference*, 2009. 2

[5] 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

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

[7] A. Malti, A. Bartoli, and T. Collins. Template-based conformal shape-from-motion from registered laparoscopic images. In *Conference on Medical Image Understanding and Analysis*, 2011. 1

[8] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua. Capturing 3D stretchable surfaces from single images in closed form. In *International Conference on Computer Vision and Pattern Recognition*, 2009. 1

[9] M. Perriollat and A. Bartoli. A computational model of bounded developable surfaces with application to image-based 3D reconstruction. *Computer Animation and Virtual Worlds*, 24(5):459–476, September 2013. 5

[10] 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, 5

[11] 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, February 2008. 1, 2

[12] 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. 1, 2, 5, 6

[13] M. Pollefeys, R. Koch, and L. van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999. 1

[14] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. In *International Conference on Computer Vision and Pattern Recognition*, 2009. 1, 2, 5

[15] 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

[16] M. Salzmann, F. Moreno-Noguer, V. Lepetit, and P. Fua. Closed-form solution to non-rigid 3D surface registration. In *European Conference on Computer Vision*, 2008. 1, 2

[17] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3D shape recovery. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1–7, August 2007. 1