Template-based Paper Reconstruction from a Single Image is Well Posed when the Rulings are Parallel

Pierluigi Taddei
Politecnico di Milano
Milano, Italy
pierluigi.taddei@polimi.it

Adrien Bartoli
LASMEA (CNRS / UBP)
Clermont-Ferrand, France
Adrien.Bartoli@gmail.com

Abstract

We deal with the 3D reconstruction of deformed paper-like surfaces given a template and a single perspective image, for which the internal camera parameters are known.

The general problem is ill-posed. We show that when the surface rulings are parallel the problem is well-posed. Given a procedure to recover the rulings direction, this particular problem is equivalent to the reconstruction of a 2D curve seen from a set of 1D camera pairs given a 1D template.

Paper can be physically modeled by exploiting local properties. This allows us to formulate the reconstruction problem by non linear variational optimization.

We provide experimental results which validate our approach on simulated and real data.

1. Introduction

Template-based monocular deformable surface reconstruction has recently received a growing interest [11]. The general case cannot be solved without prior knowledge on the observed surface [8].

For instance, in the case of single face images, 3D Morphable Models have proved to be effective to recover the camera pose and the 3D head shape [2]. Generic deformable surfaces can be modeled using triangular mesh grids and recovered by minimizing a generic regularity energy term [11]. Those methods, although effective, use empirical models to describe real deformations.

We address the case of developable surface to model material such as paper [12]. In other words, papers are considered as unstretchable surfaces with everywhere vanishing gaussian curvature. This is a realistic assumption if only smooth deformations occur. This model can be described by local constraints related to the first and second fundamental forms.

We assume that a set of point correspondences is available between the template and the (perspective) image. The problem is illustrated in Figure 1. It is formulated as a functional minimization with free boundary conditions with non linear constraints on the first and second partial derivatives of the surface. This physical model is ill-posed, in the sense that there is generally an infinite number of solutions.

We show that the reconstruction problem can be simplified to a large extent by considering those isometries that map the template to a developable surface with parallel rulings. In this case the deformed surface can be parametrized by the rulings direction and a planar generatrix curve (see Figure 2).

In this situation the reconstruction problem is well-posed: we show that this is equivalent to reconstructing a planar curve seen by 1D cameras. Our reconstruction pro-

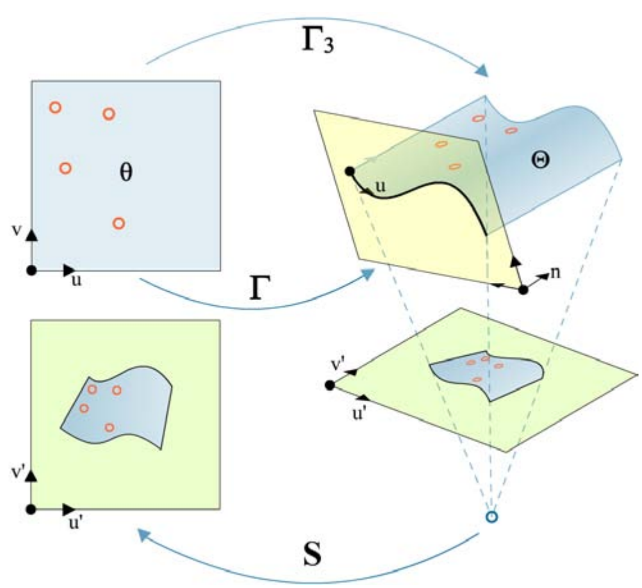

Figure 1. We tackle the template based single image paper reconstruction problem. We show that it is well-posed when the surface rulings are parallel, such as in the top right surface. We give a practical procedure for finding the generatrix and the 3D surface.

cedure has two main steps: we first reconstruct the surface generatrix plane and then a 2D curve seen from multiple 1D cameras.

Intuitively these constrained deformations (i.e. parallel rulings) correspond to bending a rectangular piece of paper by moving two opposite edges and constraining these to remain parallel. This is what happens, for instance, when book pages are deformed by keeping the binding and the opposite edge parallel.

The application range is limited. From a theoretical standpoint, however this work presents a lower bound for the well-posedness of template based reconstruction in the case of isometric deformations.

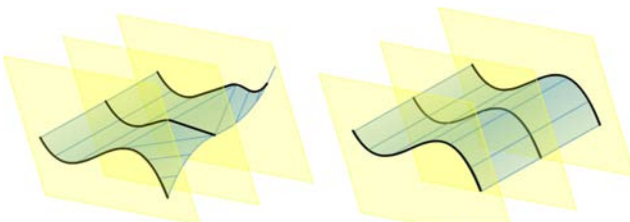

Figure 2. Two examples of paper deformations. Left: a generic isometry. Right: an isometry that allows the problem to be reduced to 1D: in this case all the rulings are parallel.

The paper is organized as follows. In Section 3 we present the formulation for the constrained two dimensional case and demonstrate in Section 4 that it is equivalent to a 2D stereo reconstruction. Section 5 describes the one dimensional problem. In Section 6 we express the constrained problem with variational optimization and present some implementation details in Section 7. In Section 8 we validate the approach by showing preliminary results on both synthetic and real data. Finally, Section 9 concludes the paper.

2. Previous Work

Paper-like surface have been successfully described using developable surfaces, which satisfy the vanishing gaussian curvature constraint. [9] proposed a quasi-minimal deformable model which showed to be effective for 3D reconstruction with more than one image. Other deformable models have been proposed by [10]. These methods, do not take advantage of the particular physical properties of the developable surfaces.

Surface reconstruction has also been performed by applying shape-from-contour mainly for document digitization. [4] assumes the pages to be generalized cylinders with straight meridians, which is equivalent to our constrained model. [6] generalizes the problem to any applicable surfaces and express it as a set of differential equations. Unfortunately shape-from-contour requires the complete knowledge of the boundary projection. Hence these methods are not applicable if the boundary is partially or fully occluded, a common case for augmentation purposes.

We exploit the variational framework. In particular, our problem has natural boundary conditions and does not have a closed form solution. Other computer vision topics which make use of variational methods either use fixed boundary conditions (such as inpainting algorithms [13, 7]) or cyclic boundary conditions for closed domains (such as level set algorithms [5, 15, 14]). In some other cases the boundaries are free to move but it is possible to recover the analytic solution to the Euler-Lagrange equations such as the Thin-Plate Spline [3].

3. Problem Statement

Given an inextensible surface with everywhere vanishing gaussian curvature, we aim to recover a function $\Gamma_3 : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ which maps a template point to the corresponding 3D point of the deformed surface. We have a perspective image of the deformed surface and we assume to know the camera internal parameters. Without loss of generality we consider the camera placed at the origin of the global reference system (see Figure 1).

Let us consider the template image. It is a compact closed region θ with boundary β. A point is represented by its coordinates $q = [u, v]^{\mathrm{T}}$. We next consider the perspective image. It represents the same closed region which has undergone an inextensible (isometric) deformation and a projection. We consider as available a set of corresponding points (q_i, q'_i).

The domain coordinates of Γ_3 will be referred as u (or u, v for the 2D case). Vectors Γ_u and Γ_{uu} represent respectively the first and second derivatives of function Γ w.r.t. the first domain coordinate u. The region $\theta \in \mathbb{R}^2$ is mapped by Γ_3 into the region $\Theta \in \mathbb{R}^3$, whereas $\beta \in \mathbb{R}^2$ is mapped by Γ_3 to the curve $B \in \mathbb{R}^3$

In the case of a generic isometry all points of the surface must obey to the following constraints:

- since the gaussian curvature is invariant under isometries and the template is a flat surface (thus with zero gaussian curvature), it must vanish.

- the metric should not change. This constraints involves that the three first fundamental form parameters are invariant under isometries.

To solve the constrained reconstruction problem, we allow only deformations that map θ to a developable surface with parallel rulings. In this case the developable surface directrix lies on a plane perpendicular to all rulings. Without loss of generality we consider in the following that π is the plane containing the directrix related to the lower edge of the template. Intuitively these deformations correspond to bending a rectangular piece of paper by moving two opposite edges and keeping them parallel. In this case, π is orthogonal to these two edges.

In particular, we show that for these deformations the problem is equivalent to a one dimensional reconstruction, where the domain θ is a one dimensional manifold, the mapping Γ is a function from $\mathbb{R}$ to $\mathbb{R}^2$ and the projection is given by a set of one dimensional cameras.

We assume that the plane π is known (refer to Section 8 for a way to recover this plane) and parametrize the mapping Γ_3 as:

$$\Gamma_3(u,v) = T \cdot \tilde{\Gamma}(u,v) = T \cdot \begin{pmatrix} \Gamma(u) \\ v \end{pmatrix} \quad (1)$$

where $\Gamma : \mathbb{R} \rightarrow \mathbb{R}^2$ maps the horizontal coordinate of the template onto a curve on π equipped with a suitable coordinate system (x, y, n_π) and T is the transformation that maps this latter coordinate system to the global one (X, Y, Z). Notice that T might be considered fixed if π is known in advance and the reference system (x, y, n_π) is defined.

This representation of the surface Θ allows one to disregard the gaussian curvature constraint since it is enforced by construction. The surface is finally projected onto the image plane by the known camera matrix $S \sim K \cdot \begin{pmatrix} I & 0 \end{pmatrix}$ where $\sim$ means equality up to scale.

The following section demonstrates that the problem at hand reduces to a one dimensional formulation that exploits multiple 1D cameras

4. A One Dimensional Formulation

Let as consider a point $q \in \theta$. This point is mapped by Γ_3 to the 3D space and then projected by the camera to:

$$q' \sim S \cdot T \begin{pmatrix} \tilde{\Gamma}(u,v) \\ 1 \end{pmatrix} = S_T \begin{pmatrix} \Gamma(u) \\ v \\ 1 \end{pmatrix}. \quad (2)$$

We write the columns of the projection matrix $S_T = S \cdot T$ as $S_T = \begin{pmatrix} s_1 & s_2 & s_3 & s_4 \end{pmatrix}$ and rewrite Equation (2) as:

$$q' \sim \begin{pmatrix} s_1 & s_2 & (s_4 + s_3 v) \end{pmatrix} \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix} = S_v \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix}. \quad (3)$$

By separating the two components of q' we get:

$$q'_u \sim S_x \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix} = \begin{pmatrix} S_{va} \\ S_{vc} \end{pmatrix} \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix} \quad (4)$$

$$q'_v \sim S_y \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix} = \begin{pmatrix} S_{vb} \\ S_{vc} \end{pmatrix} \cdot \begin{pmatrix} \Gamma(u) \\ 1 \end{pmatrix}, \quad (5)$$

where S_{va}, S_{vb}, S_{vc} are the three row vectors of matrix S_v. Each of these two equations represents a 1D projection of the same point $\Gamma(u)$ onto a virtual 1D camera projection line. The projection matrices S_x and S_y in particular linearly depend on coordinate v. This means that the constrained 2D case is equivalent to the 1D reconstruction problem that exploits two pairs of 1D cameras for each surface section orthogonal to the rulings.

5. Formulating the Constraints

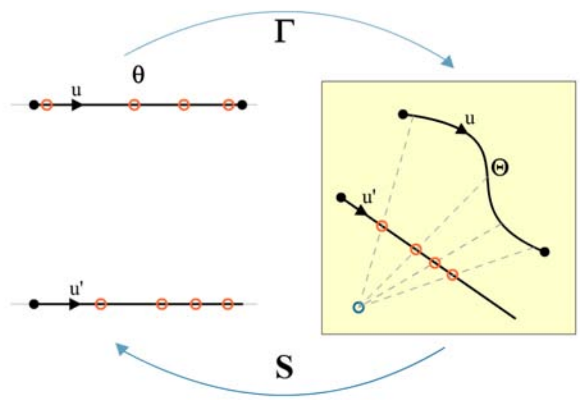

Figure 3. The one dimensional equivalent problem. In this case Γ is represented by a curve on the plane and the projection is done by 1D cameras such as S.

The equivalent problem for a one dimensional manifold domain is stated as follows (see Figure 3). We consider a template consisting of a straight line θ. A point on the template is represented by its coordinate u (and thus the domain is equipped with a unitary metric operator). The template is isometrically mapped by Γ to a planar curve Θ. We have a set of 1D projections of Θ given by known 1D cameras (2×3 matrices). For each known camera there is a different set of point correspondences between the template and the projected Θ projection: (u, u') where $u, u' \in \mathbb{R}$. The problem is thus to recover Γ such that:

- the reprojection error of the point correspondences is minimized
- the deformation induced by Γ is isometric, thereby preserving lengths along the curve w.r.t. the template

Exploiting only these two constraints the problem is well-posed as far as the extracted points are concerned, but there are still multiple solutions for all other Θ points. In particular consider two subsequent correspondence points. If their euclidean distance in 2D is smaller than their geodesic distance (the distance on the template) then there exist an infinite number of curve segments connecting the two points which all share the same length.

The introduction of a smoothing term, which must be minimized, allows one to recover a solution with, for instance, the least curvature. This is equivalent to finding the smoothest surface.

The isometry constraint is expressed by considering the metric induced by Γ, which depends on the first derivative, and in particular by stating that:

$$\|\Gamma_u\|^2 = 1 \quad \forall u \in \Theta \tag{6}$$

.

The problem then can be expressed as a functional minimization that depends on the first and second derivatives of Γ, since all the constraints are represented by local properties.

6. Variational Formulation

In order to solve the constrained 2D problem we express it as the following functional minimization:

$$\Gamma = \arg\min_{\Gamma} \left(E_d[\Gamma] + E_m[\Gamma_u] + E_s[\Gamma_{uu}] \right), \tag{7}$$

where:

- $E_d[\Gamma]$ represents the *data term* which states that the reprojection $r(q, \Gamma, S, T)$ of the feature point q lying on the surface Γ should be as close as possible to the corresponding image feature q'. r is the reprojection function, which depends on the projection matrix S and the surface Γ and has the form:

$$r(q, \Gamma, S, T) = \frac{S_{ab} \cdot T \cdot \begin{pmatrix} \Gamma(u) \\ v \end{pmatrix}}{S_c \cdot T \cdot \begin{pmatrix} \Gamma(u) \\ v \end{pmatrix}}, \tag{8}$$

  where S_{ab} is the matrix composed by the first two rows of S and S_c is the last row of S;

- $E_m[\Gamma_u]$ represents the *metric term* which is used to enforce the surface deformation to be an isometry;

- $E_s[\Gamma_{uu}]$ represents the *smoothing term* which states that the second partial derivatives of the surface should be as small as possible, describing the fact that we assume the deformation to be smooth.

With this notation we write the energy terms as:

$$E_d[\Gamma] = \sum_i \|r(q_i, \Gamma, S, T) - q_i'\|^2 \tag{9}$$

$$E_m[\Gamma_u] = \int_\theta \left(\|\Gamma_u(u)\|^2 - 1 \right)^2 \tag{10}$$

$$E_s[\Gamma_{uu}] = \int_\theta \|\Gamma_{uu}(u)\|^2 \tag{11}$$

Equation (9) may be restated as an integral on the domain θ as:

$$E_d[\Gamma] = \int_\theta \left(\sum_i G_{i,\sigma}(q) \, \|r(q_i, \Gamma, S) - q_i'\|^2 \right) dq \tag{12}$$

where $G_{i,\sigma}(q)$ represents a gaussian function centered on feature q_i of the template and standard deviation σ. This particular form of the data term allows one to express the integrand as a continuous functional over the domain.

The variational problem is thus of the form:

$$\arg\min_{\Gamma} \left(E[\Gamma] \right) = \arg\min_{\Gamma} \left(e(u, \Gamma, \Gamma_u, \Gamma_{uu}) \right) \tag{13}$$

The integrand function e depends on Γ and its first and second derivatives. By applying the variational fundamental theorem we obtain two partial differential equations depending on Γ and its derivatives (up to the fourth). These equations represent a set of constraints on Γ in order to be a local minimum of the functional and must be satisfied in each point of the domain θ. We refer the reader to [14, 1] for a good description of the variational framework.

Notice that our problem cannot make use of fixed boundary conditions since the two end points must be able to move freely. This means that we have to subject these to particular natural boundary conditions which are automatically given by the application of the fundamental theorem.

7. Implementation Details

To solve the set of Euler-Lagrange equations on each point of θ and the natural boundary condition on each point of β we parametrize the function Γ over the domain with a set of N pairs of unknown (x_i, y_i). To correctly evaluate its derivatives we add a set of extra nodes outside θ. Notice that we do not enforce the Euler-Lagrange equations on these nodes, although we use them to evaluate the equations for domain points close to the boundary. The final system has $2(N+8)$ non linear equations over $2(N+8)$ unknowns that we solve using standard Gauss-Newton descent.

The lack of fixed or cyclic boundary conditions poses problems w.r.t. the approximation of the derivatives. The central scheme, although symmetric, gives rise to solutions with loosely connected nodes which cannot be avoided. The forward and backward schemes, on the other hand, limit this sampling effect but are asymmetrical and have slower convergence. In order to maintain a symmetric scheme and, at the same time, avoid these unwanted effects we double all equations using the forward and the backward approximations.

The three energy terms are weighted using a set of fixed parameters, whose values are empirically selected. In particular, we run subsequent minimizations by having at first an higher data term weight in order to reach a solution that minimizes the reprojection error and then another with an higher metric term weight.

8. Experimental Validation

Solving the 2D constrained problem requires to know the directrix plane π, since the optimization is done on a curve lying on this plane. If the projection of the four corners of the template and the length of one of the sides are known, π can be exactly recovered. This is a problem with four unknowns (the perspective depth of the four corner) which can be solved by applying the constraints of coplanarity and orthogonality among the four segments connecting the extracted points and the known length of one of the edges. We solve this non linear problem using a least-squares procedure that generally gives a good estimate of the plane. In the case where the corners are not visible it is still possible to exploit the image of four points lying on two different rulings (as has been done for the surface in Figure 8). If these points are not available one can exploit some deformable warp between the template and the image, recovered from the point correspondences. The warp is used to detect the largest template rectangle visible in the perspective image. The four extrapolated corners are then used to recover the plane.

The known template dimensions are used to obtain a rectified and properly scaled template from an initial perspective image.

Our method has been validated on both synthetic and real data. We performed simulated experiments to recover the pose of the surface and to perform comparisons with ground truth. We show a sample reconstruction result for these sets of experiments in Figure 4 and Figure 5. In all the experiments the initial solution is given by a flat surface orthogonal to plane π. The results where good even in the presence of gaussian noise on the initial points coordinate values (we used a maximum standard deviation $\sigma = l/250$ where l is the horizontal image plane length).

We performed preliminary real experiments that showed promising. We show some of the results in Figure 7 and Figure 8 where the solutions have been recovered exploiting around 30 correspondences (Figure 6). In all cases the reconstruction we obtained was realistic. We performed the augmentation of the image by replacing the original template with a custom image or by adding virtual 3D content to the scene (Figure 7). In the second experiment, the surface is only partially visible in the image. Nevertheless it has been fully reconstructed (Figure 9).

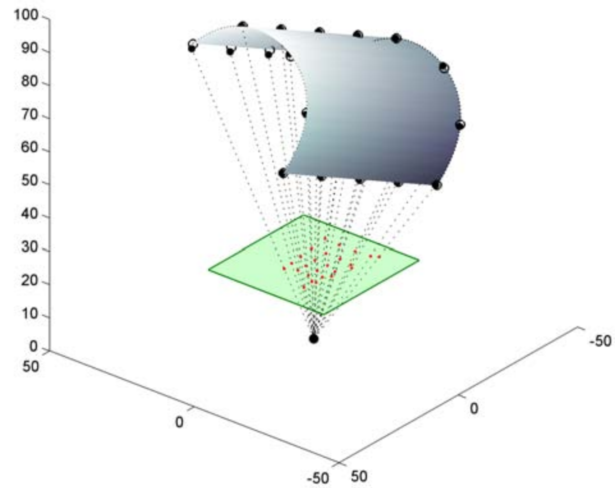

Figure 4. Final solution for the surface reconstruction on simulated data. For this experiment we exploited 25 feature point correspondences to recover the pose of the surface.

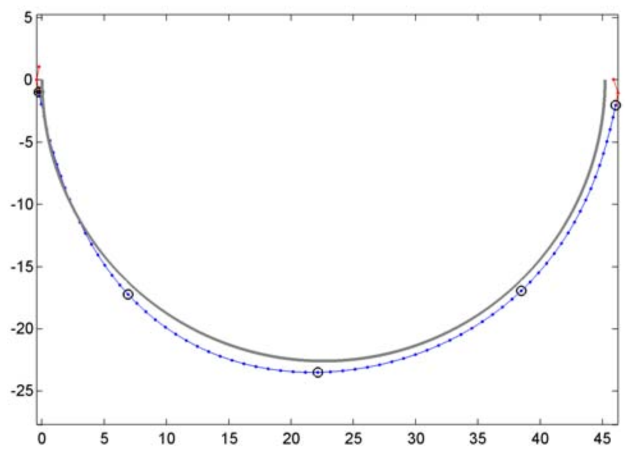

Figure 5. Comparison of the final surface directrix (blue dotted line) with respect to the ground truth (gray line). The circled nodes are related to the peaks of the data term gaussians. The red points represent the extra nodes used for the discrete approximations.

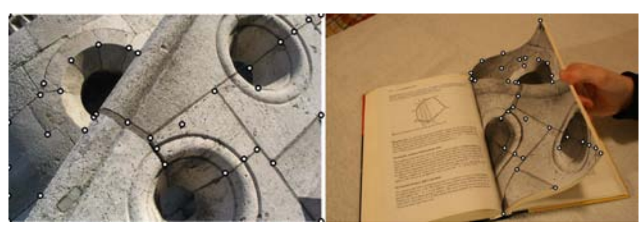

Figure 6. Point correspondences between the template and the image for the first experiment.

9. Conclusion

We presented a method to reconstruct deformable *paper-like* surfaces when these are subject to a subset of the possible isometries. In particular we dealt with the case of surfaces with parallel rulings. We showed that the problem can be cast to a 2D stereo reconstruction of the generatrix with multiple 1D cameras. This demonstrates that the problem is well-posed.

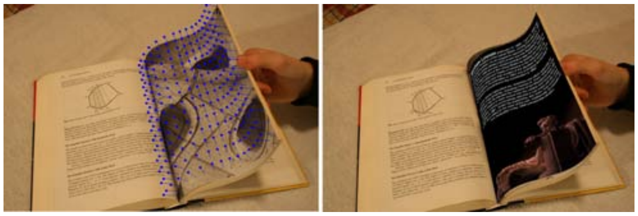

Figure 7. Left: the solution is illustrated by projecting the surface normals on the original view. Right: the same image has been retextured.

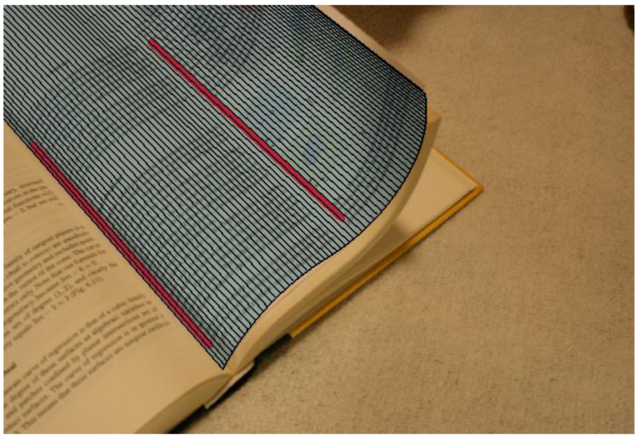

Figure 8. The image of the second experiment. In this case the paper is partially occluded. The projection of the reconstructed surface is shown in blue. To detect the generatrix plane we exploited the two segments shown in red.

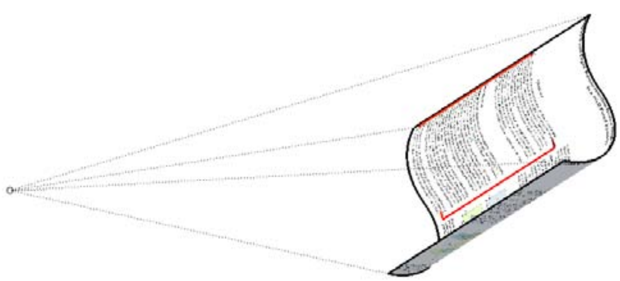

Figure 9. A synthetic view of the reconstructed surface.

We are currently investigating whether more generic isometries are well-posed problems, such as the case in which only the two paper edges are constrained to be parallel. In practical applications the initial directrix plane estimation influences the overall accuracy of the reconstruction. This will be investigated in future work. The variational approach could be directly used to solve the general ill-posed problem. In this situation the formulation becomes more complex because we must add terms related to the gaussian curvature and to the metric constraints. In addition, the standard numerical minimization becomes computationally intensive since both the number of unknowns and the number of equations drastically increase.

Acknowledgments The authors would like to thanks Mathieu Perriollat for the useful discussions.

References

[1] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855–867, 1990.

[2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *CGIT*, 1999.

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

[4] F. Courteille, J.-D. Durou, and P. Gurdjos. Shape from contour for the digitization of curved documents. In *ACCV*, 2007.

[5] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *ICCV*, 2007.

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

[7] F. Lauze and M. Nielsen. A variational algorithm for motion compensated inpainting. In *BMVC*, 2004.

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

[9] M. Perriollat and A. Bartoli. A quasi-minimal model for paper-like surfaces. In *IEEE/ISPRS Workshop BenCOS*, 2007.

[10] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. In *CVPR*, 2005.

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

[12] H. Pottman and J. Wallner. *Computational Line Geometry*. Springer, 2001.

[13] T. Shih and R. Chang. Digital inpainting survey and multilayer image inpainting algorithms. In *ICITA*, 2005.

[14] J. Solem. *Variational problems and level set methods in computer vision: theory and applications*. PhD thesis, Lund University, 2006.

[15] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *ICCV*, 2003.