A Pixel-Based Approach to Template-Based Monocular 3D Reconstruction of Deformable Surfaces

Abed Malti   Adrien Bartoli   Toby Collins
ALCoV-ISIT, Université d'Auvergne
Clermont-Ferrand, France.
http://isit.u-clermont1.fr/

Abstract

Most of the previous work on template-based deformable 3D surface reconstruction using a single view is feature-based. We propose a pixel-based formulation in a variational framework; the unknown is the surface function. The color discrepancy between the template and the deformed images is formalized as a functional of the surface function. The main difficulty in such a formulation arises when the surface self-occludes which induces discontinuities in the discrepancy measure at the self-occlusion boundary. Based on previous work on 3D rigid surface reconstruction, we rigorously formalize the visibility as a continuous functional of the surface function. It is derived in the template for visible/self-occluded regions in the deformed image. The gradient of the color discrepancy is computed with respect to the surface function. The minimization smoothly updates the surface function to fit the self-occlusion boundary. Gradient descent is initialized from feature-based 3D reconstruction.

Our experimental results on simulated and real data show that during the minimization of the color discrepancy, the self-occlusion boundary of the reconstructed surface moves to its correct location in the image. We show quantitatively that in the template image, the accuracy of visible/self-occluded areas is improved to a significant extent.

1. Introduction

Monocular 3D reconstruction of non-rigid surfaces is an important task for many applications such as augmented reality. This is a difficult problem since the appearance of an imaged surfaces varies due to several phenomena such as camera pose, surface deformation, lighting, motion blur and self-occlusions. During the past few years, feature-based reconstruction methods have been proven efficient to recover the 3D shape of deformable surfaces [3, 12]. However, since these methods rely on sparse feature matches
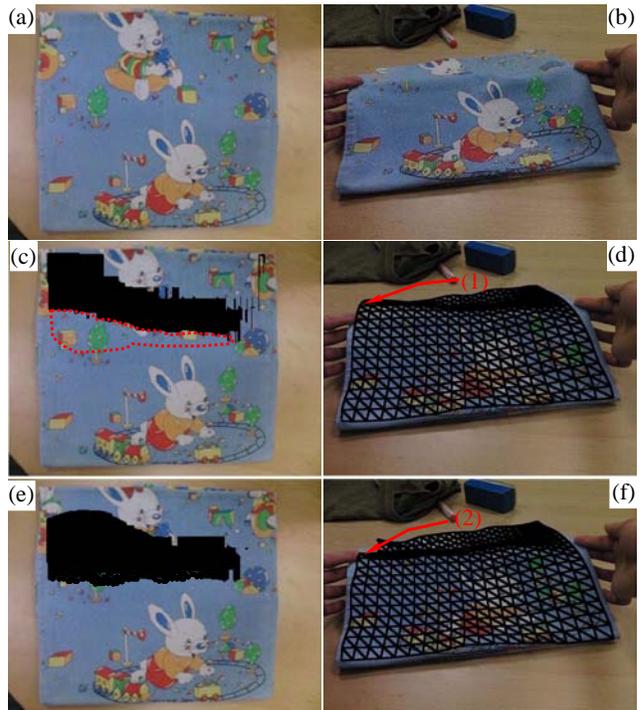

Figure 1. (a) Template surface. (b) Deformed surface. Second row is for feature-based method [3]. Third row is for our pixel-based method. In (c) and (e) self-occluded area in black. The arrow (2) in (f) shows that the self-occlusion boundary is better fitted with the proposed pixel-based method than with a feature-based method as is pointed out by arrow (1) in (d). Thanks to the proposed method, the missed self-occluded area illustrated in dotted red in (c) is recovered as shown in (e).

to reconstruct the deformed surface they may provide a rough reconstruction when data are missing in some critical regions of the deformation. For instance, in case of a self-occlusion, the 3D reconstruction may be inaccurate around the self-occlusion boundary, as shown in figure 1.d. In pixel-based 3D rigid surface reconstruction methods the surface is more densely covered by using image intensities

over all the visible pixels. Based on an initial estimate of the surface function between the template and deformed images, the color discrepancy is minimized [2]. Usually this minimization involves computing the gradient of the discrepancy functional. However, in the presence of a self-occlusion, a simple binary function to represent the visibility will not be continuous at the boundary between the visible and the self-occluded regions (the visibility function moves abruptly from 1 to 0). Some authors bypass this problem by learning beforehand a set of statistical 3D models of the deformed shape [10]. Moreover, it is obvious that the visibility function depends on the surface function and a basic binary function does not take into account this dependency. Based on work by Solem *et al.* [14] on the geometric formulation of gradient descent for variational problems with moving surfaces, Gargallo *et al.* [5] derived a 3D surface reconstruction algorithm. In their method, the visibility function is explicitly defined with respect to a rigid 3D surface as being the zero-set of a level-set function. The gradient descent algorithm is initialized using a rough estimate of the rigid surface from rigid Structure-from-Motion. In the context of template-based 3D reconstruction of deformable surfaces from one single view, the application of such a method is not straightforward. The difficulty lies in the formulation of the visibility function with respect to the surface function.

In this work, the first pixel-based approach to 3D reconstruction of a deformable surface using one single view is proposed. Based on methods for 3D rigid surface reconstruction, it expresses in a variational framework the color discrepancy with respect to the surface function. The visibility is defined as a continuous functional from the template to the deformed surface. The derivative of the color discrepancy is computed in the template with respect to the surface function which facilitates gradient descent minimization. The 3D reconstructed surface is given in the camera frame of the deformed image.

Paper organization. Section 2 gives related work. Section 3 presents basic definitions and concepts. Section 4 presents the proposed pixel-based variational method. Section 5 reports our experimental results and finally section 6 concludes.

Notation. Vectors are in bold (*e.g.* $\mathbf{q}$). Matrices are in uppercase sans-serif fonts (*e.g.* Π). The identity matrix is denoted $\mathbf{I}$. $||.||$ denotes the vector two norm in a real vector space of finite dimension. Images are denoted using calligraphic fonts (*e.g.* $\mathcal{I}$). They are seen as functions from a compact domain $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}^3$ for the three RGB color channels . For instance $\mathcal{I}(\mathbf{q})$ is the image value of pixel $\mathbf{q} \in \Omega$. Bilinear interpolation is used for sub-pixel coordinates. The gradient of a scalar valued func-

tion f with respect to vector $\mathbf{x} = (x_1 \cdots x_n)^\top$, is denoted $\nabla_{\mathbf{x}} = (\partial f x_1 \cdots \partial f x_n)$. H designate the Heaviside step function valued to zero for real negative arguments and to one for real positive arguments. δ stands for the Dirac impulse valued to one at zero and to zero elsewhere. The abbreviation *iff* states for the logical equivalence (if and only if).

2. Related work

Up to now, only 2D warp estimation has been addressed using pixel-based approaches to deformable surfaces with known template [1]. Even recent works has addressed the problem of self-occlusion in 2D warp estimation [11, 6], to the best of our knowledge there is no existing pixel-based method to template-based 3D reconstruction of deformable surfaces. However, the problem we tackle in this paper has been addressed separately in the computer vision community, where in one hand several algorithms have been proposed for feature-based (either sparse or even some times using denser matching methods like template-matching) monocular 3D reconstruction of non-rigid shapes and in the other hand a set of pixel-based approaches have been derived for 3D reconstruction of rigid surfaces with self-occlusion handling. In feature-based methods, the deformed surface is obtained by measuring distances between sparse features matched between a template and a target image. These approaches may differ by the type of representation of the surface. The point-wise methods use a sparse representation of the 3D surface [9], while other methods make use of complex surface models like triangular meshes [12] or smooth parametric maps [3, 8, 15]. Also they may differ by the physical assumptions on the surface; some of them use analytic constraint in the case of isometric deformation [3] while others learn local statistical models of deformation [13] relying on template matching which enables denser coverage of the surface. In [7] sparse feature correspondences are combined with silouhette points to fit an implicit 3D mesh model to a given surface deformation. The purpose of this method is similar to the goal of the proposed approach since both handle self-occlusions. However, they differ in the way the visibility is handled; while in [7] silouhettes are taken to be self-occlusion border, in our approach we rigorously formalize the visibility functional with respect to the deformable surface in a variational framework which allows us to derive continuous functional expressions and use gradient descent methods.

In pixel-based approaches to 3D reconstruction of rigid shapes, developments of the last decade have led to the formulation of many interesting problems in a variational setting. Surface evolution is one of the most successful approaches in 3D geometric reconstruction [4, 17, 14]. The idea of starting with an initial rough approximation of the surface and then deform it so that it improves some color

discrepancy score has been proved to be efficient. However, these methods describe the visibility function as binary function (*i.e.* one for visible regions and zero for self-occluded ones) which imposes to use some computational tricks to overcome the discontinuity in the self-occlusion boundary. Tsai *et al.* [16] have defined the visibility, using level-set functions, as a continuous function when the view point and the scene objects move. Recently, based on works by [16] and [14], Gargallo *et al.* [5] propose continuous expression of the color discrepancy in the self-occlusion boundary to successfully produce 3D reconstruction of rigid surfaces using multiple views. In our work, we extend this approach to template-based 3D reconstruction of deformable surface using one single view.

3. Background and definitions

A **template image** is a compact domain $\Omega \subset \mathbb{R}^2$ for which the 3D shape of the **template** surface Γ is known. A **surface function** is a $\mathcal{C}^k$-function φ which maps the compact domain Ω to the deformed surface Γ' embedded in $\mathbb{R}^3$:

$$\Omega \to \mathbb{R}^3 : \mathbf{q} \mapsto \mathbf{Q}' = \varphi(\mathbf{q}) \tag{1}$$

As it represents a regular surface, the integer k is constrained to be at least two. A **target image** is a compact domain $\Omega' \subset \mathbb{R}^2$ which contains the visible projected points $\mathbf{q}' \in \Omega'$ of the deformation of the surface points $\mathbf{Q}' \in \Gamma'$ given a point of view:

$$\Pi \begin{pmatrix} \mathbf{Q}' \\ 1 \end{pmatrix} = \mu \begin{pmatrix} \mathbf{q}' \\ 1 \end{pmatrix}, \ \mathbf{q}' \in \Omega' \tag{2}$$

where $\Pi \in \mathbb{R}^{3 \times 4}$ is the projection matrix which encodes both camera intrinsics and extrinsics. μ refers to the depth of the 3D point $\mathbf{Q}'$. Without loss of generality Π can be assumed as being the canonical projection with known camera calibration parameters. To simplify the notation, we refer to $\mathcal{W} : \mathbf{q} \mapsto \mathbf{q}'$ as the 2D warp which concatenates equations (2) and (1). This allows to define $\Omega^* \subset \Omega'$ as being the mapping of Ω by the warp $\mathcal{W}$.

4. Pixel-Based Variational Formulation

4.1. First Variational Formulation

A geometric pixel-based 3D reconstruction problem can be formulated in a variational framework as the minimization of the photometric error between the template and the target images:

$$\min_{\varphi} \quad \mathcal{E}_{\mathcal{I}} + \lambda_p \mathcal{E}_p \tag{3}$$

where:

$$\mathcal{E}_{\mathcal{I}} = \int_{\mathbf{q} \in \Omega} \| \left(\mathcal{I}'(\mathcal{W}(\mathbf{q})) - \mathcal{I}(\mathbf{q}) \right) \nu_\Gamma(\mathbf{q}, \varphi) \|^2 \, d\mathbf{q} \tag{4}$$

represents the color discrepancy weighted by the visibility function ν_Γ defined in the template domain by:

$$\nu_\Gamma : \begin{array}{rcl} \Omega & \to & \{1, 0\} \\ \mathbf{q} & \mapsto & \begin{cases} 1 & \text{if } \mathcal{W}(\mathbf{q}) \text{ is visible,} \\ 0 & \text{otherwise.} \end{cases} \end{array} \tag{5}$$

The term $\mathcal{E}_p$ accounts for some prior knowledge about the physical constraints of the deformable surface (*e.g.* extensiblity, smoothness, etc). λ_p is real positive number to tune the importance of this prior over the photometric error.

4.2. Visibility in the template space

It is obvious that the visibility function depends on the surface posture and the viewing point which are implicitly given in the target image. However, as it is described this dependency is not explicit. It has to be exhibited for gradient computation purposes to deform the surface according to the visible and the self-occluded areas. For this reason, it has been shown [16, 14] that the visibility function, $\nu_\Gamma(\mathbf{q}, \varphi)$, of a regular surface observed from a vantage point depends on the level-set function that is defined as the zero-set of this surface. Indeed, since the deformed surface Γ is a regular surface of codimension 1 in $\mathbb{R}^3$, it can be represented as the zero set of a $\mathcal{C}^2$-function in $\mathbb{R}^3$, *i.e.* $\Gamma = \{(x, y, z) \in \mathbb{R}^3 | \phi(x, y, z) = 0\}$. The sets $\{(x, y, z) \in \mathbb{R}^3 | \phi(x, y, z) < 0\}$ and $\{(x, y, z) \in \mathbb{R}^3 | \phi(x, y, z) > 0\}$ are called the inside and the oustide of ϕ. From the vantage point, these sets are respectively behind and before of the surface Γ, see figure 2. As expected, it is this property that allows us to well define the visibility as being dependent of the associated level-set function. One of our contributions is to exhibit the visibility function in the template space and to compute its gradient with respect to surface deformation so that it can be taken into account in the minimization of the photometric error. For this purpose, we define a level-set function as a functional in the template domain:

$$\phi_\varphi : \begin{array}{rcl} \Omega \times \mathbb{R} & \to & \mathbb{R} \\ (\mathbf{q}, \lambda) & \mapsto & \phi_\varphi(\lambda \varphi(\mathbf{q})) \end{array} \tag{6}$$

where λ is a real positive parameter. $\lambda \varphi(\mathbf{q})$ is the sight line (joining the view point and the surface point $\varphi(\mathbf{q})$) which intersects the different level sets of ϕ_φ. According to this parametrization, the zero-set becomes $\Gamma = \{\mathbf{q} \in \Omega | \phi_\varphi(\mathbf{q}, 1) = 0\}$. If there is no self-occluded area, the exterior set (behind Γ) is $\{\mathbf{q} \in \Omega, \lambda > 1 : \phi_\varphi(\mathbf{q}, \lambda) < 0\}$ and the interior set (before Γ) is $\{\mathbf{q} \in \Omega, \lambda < 1 \ \phi_\varphi(\mathbf{q}, \lambda) > 0\}$. If there is some self-occluded area, the parameterized line $\lambda \varphi(\mathbf{q})$ may cross a level-set curve at least twice. This property will help us to determine the self-occluded region as described below.

Thanks to the exterior and interior sets, the visibility of any surface point $\varphi(\mathbf{q}) \in \Gamma$ can be determined from the
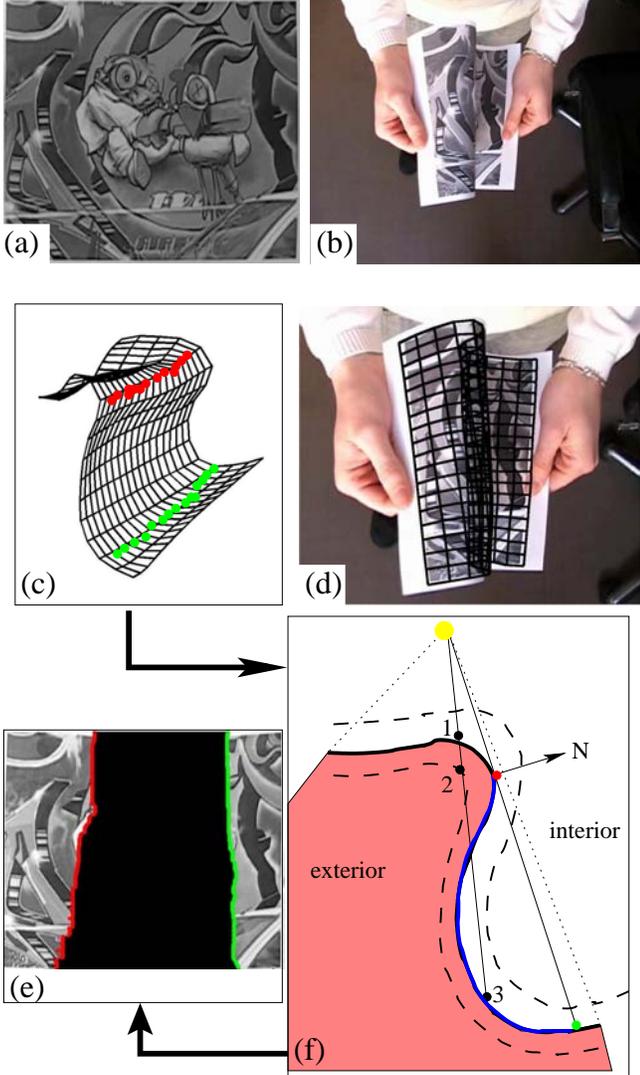
Figure 2. Visibility/Occlusion sets. (a) represents a template image. (b) represents the target image where the deformed surface self-occludes. (c) and (d) show the computed 3D model of the deformed surface. (f) sketches how the visibility/occlusion sets are determined. The red point on the surface represents a point of the horizon. The green point on the surface represents the terminator point which is the dual of the red one. Point 1 is the self-occluder of point 3. Point 2 is the result of the mapping by β of point 3.

values that ϕ_φ takes when λ varies from 0 to 1. For a given $\mathbf{q} \in \Omega$ if all the values $\phi_\varphi(\mathbf{q}, \lambda)$ are positive for all $0 \leq \lambda \leq 1$ then the line of sight from the view point to $\varphi(\mathbf{q})$ is in the interior of Γ, which means that $\mathbf{q}$ is visible. However, if any of the level-sets values $\phi_\varphi(\mathbf{q}, \lambda)$ is negative, which means that the line of sight has crossed the surface Γ at least two times and has passed through the exterior before hitting the point $\mathbf{q}$ to which corresponds $\lambda = 1$. In this case it is obvious that $\varphi(\mathbf{q})$ is self-occluded and its self-occluder on the surface is the first point of the intersection of the sight

line $\lambda\varphi(\mathbf{q})$ with the surface. It is defined as:

$$\lambda_\mathbf{q} = \beta(\mathbf{q}, \lambda) = \operatorname*{argmin}_{\lambda \in \mathbb{R}}\{|\phi_\varphi(\mathbf{q}, \lambda)|\} \qquad (7)$$

Tsai *et al.* [16] have shown that β is a continuous function on $\Omega \times \mathbb{R}$ (in their notation, it is continuous in the 3D space and since the functional φ is continuous in $\mathbf{q}$, by the composition of continuous functions the result holds). If $\mathbf{q} \in \Omega$ is visible then $\phi_\varphi(\mathbf{q}, \lambda_\mathbf{q}) = 0$ and if is self-occluded then $\phi_\varphi(\mathbf{q}, \lambda_\mathbf{q}) < 0$. Henceforth, visible and self-occluded points can be characterized by the following proposition:

Proposition 4.1 *A template point* $\mathbf{q} \in \Omega$ *is said to be **visible** iff* $(\phi_\varphi(\mathbf{q}, \lambda_\mathbf{q}) = 0$, *or equivalently, iff* $\beta(\mathbf{q}, \lambda) = 1$. *However a template point* $\mathbf{q} \in \Omega$ *is said to be **occluded** iff* $(\phi_\varphi(\mathbf{q}, \lambda_\mathbf{q}) < 0$, *or equivalently, iff* $\beta(\mathbf{q}, \lambda) < 1$. *Moreover,* $\mathbf{q}_o$, *such that* $\mathbf{q}_o = \beta(\mathbf{q}, \lambda)\varphi(\mathbf{q})$, *is said to be the **occluder** of* $\mathbf{q}$.

The set of all visible points is denoted $\mathcal{V} \in \Omega$ and the set of all self-occluded points is denoted $\mathcal{O} \in \Omega$. Among the pairs self-occluder/occluded points, there is a pair of subset of particular interest which are called the **horizon** and the **terminator**. They self-occur when the sight line that joins the origin to the self-occluded point is tangent to the surface Γ at the self-occluder point:

Definition 4.1 *A visible point* $\mathbf{q} \in \mathcal{V}$ *is said to be on the* **horizon** $\mathcal{H}$ *iff it satisfies the two following conditions:*

1. $\mathbf{N}_\Gamma^\top(\mathbf{q})\,\varphi(\mathbf{q}) = 0$

2. $\mathbf{q}$ *is an self-occluder*

where $\mathbf{N}_\Gamma(\mathbf{q})$ *states for the normal of* Γ *at the point* $\varphi(\mathbf{q})$. *An self-occluded point is said to be on the **terminator** $\mathcal{T}$ iff its self-occluder is on the horizon.*

It can be easily seen that condition 1 of this definition is necessary but not sufficient since in the case where condition (2) is not fulfilled, then the point self-occludes a point outside Γ (a background point) which does not correspond to a self-occlusion. Using the warp $\mathcal{W}$, we can write:

$$\mathcal{W}(\mathcal{V}) \cap \mathcal{W}(\mathcal{O}) = \mathcal{W}(\mathcal{H}) = \mathcal{W}(\mathcal{T}) \qquad (8)$$

This means that in the target image horizon and terminator images are superposed. The relation is even stronger, since to every point $\mathbf{q} \in \mathcal{H}$ corresponds its dual point $\mathbf{q}' \in \mathcal{T}$. Dual points have the same projection in the target image:

$$\mathcal{W}(\mathbf{q}) = \mathcal{W}(\mathbf{q}'), \text{ for all } \mathbf{q} \in \mathcal{H} \text{ and its dual } \mathbf{q}' \in \mathcal{T} \quad (9)$$

This property defines a one-to-one correspondence between the horizon and the terminator points:

$$\eta : \mathcal{H} \to \mathcal{T} \qquad (10)$$

According to these developments, it follows that $\phi_\varphi \circ \beta$ is an implicit representation of the self-occluded surface and it can be linked to the visibility function as:

$$\nu_\Gamma(\mathbf{q}, \varphi) = H(\phi_\varphi(\mathbf{q}, \beta(\mathbf{q}, \lambda))) \qquad (11)$$

In order to compute gradient expression of the visibility function, the functional $\nu_\Gamma = H \circ \phi_\varphi \circ \beta$ is considered as a distribution since the Heaviside function is not continuous.

4.3. Gradient of the Data functional

We assume a variation of the function φ which produces an evolution of the surface Γ that does not change the topology of the level sets. Henceforth, the variation of the level set can be described as $\phi_{\varphi_1} = \phi_\varphi + (\varphi_1 - \varphi)^\top \psi$. In this case, the visibility sets change in the template and the differential of the visibility function is given by the chain rule in the distribution space as:

$$\nabla_\varphi \nu_\Gamma(\mathbf{q}, \varphi) = \delta(\phi_\varphi(\mathbf{q}, \beta(\mathbf{q}, \lambda))) \, \psi(\mathbf{q}, \beta(\mathbf{q}, \lambda)) \qquad (12)$$

where δ is the Dirac function and ∇_φ denotes the gradient with respect to the function φ. To make it explicit, one needs to instantiate the function φ in a given basis function. This part of the computation is addressed in the next section. ψ is the variation of the level set function when the surface is deformed. Then, if we consider the photometric difference as a functional of the deformable surface:

$$g(\mathbf{q}, \varphi) = \mathcal{I}'(\mathcal{W}(\mathbf{q})) - \mathcal{I}(\mathbf{q}) \qquad (13)$$

Its gradient in the visible part of Ω is:

$$\nabla_\varphi \left(g(\mathbf{q}, \varphi) \nu_\Gamma(\mathbf{q}, \varphi) \right) = \nabla_\varphi \mathcal{W}(\mathbf{q}, \varphi) \nu_\Gamma(\mathbf{q}, \varphi) + \\ g(\mathbf{q}, \varphi) \delta(\phi_\varphi(\mathbf{q}, \beta(\mathbf{q}, \lambda))) \psi(\mathbf{q}, \beta(\mathbf{q}, \lambda)) \qquad (14)$$

Notice that the second term of the right hand side would not appear without an implicit formulation of the visibility function. This term supports the surface evolution by considering evolution in the level set function. However since the δ distribution is not a smooth function, it brings non-feasible deformation and it has to be approximated by an admissible function to allow admissible surface deformation. In practice, this is done simply by approximating it with a narrow gaussian.

4.4. Second Variational Formulation

Taking into account the properties related to the visibility and self-occluded regions of a template space Ω, the pixel-based 3D reconstruction is then formulated as follows:

$$\min_\varphi \quad \mathcal{E}_\mathcal{I} + \lambda_n \mathcal{E}_n + \lambda_{\mathcal{H},\mathcal{T}} \mathcal{E}_{\mathcal{H},\mathcal{T}} + \lambda_p \mathcal{E}_p \qquad (15)$$

with:

- $\mathcal{E}_\mathcal{I}$ the data error as defined in (4) with the visibility function as defined in (11).

- $\mathcal{E}_n$ is the orthogonal constraint on the horizon points:

$$\mathcal{E}_n = \int_{\mathbf{q} \in \mathcal{H}} \| \left(\mathbf{N}_\Gamma(\mathbf{q}) \varphi(\mathbf{q}) \right) \|^2 \, d\mathbf{q} \qquad (16)$$

- $\mathcal{E}_{\mathcal{H},\mathcal{T}}$ expresses the duality property on the horizon/terminator points:

$$\mathcal{E}_{\mathcal{H},\mathcal{T}} = \int_{\mathbf{q} \in \mathcal{H}} \| \left(\mathcal{W}(\mathbf{q}) - \mathcal{W}(\eta(\mathbf{q})) \right) \|^2 \, d\mathbf{q} \qquad (17)$$

  where η is the one-to-one mapping between horizon and terminator sets as defined in (10). This constraint ensures that the projection of the horizon and the terminator points coincide in the image.

- $\mathcal{E}_p$ is as defined in the first formulation and λ_n, $\lambda_{\mathcal{H},\mathcal{T}}$, λ_p are real positive weights.

4.5. Algorithm Description

The proposed pixel-based variational method to 3D reconstruction of deformable surface is implemented as follows:

Step 1: The 3D surface can be initialized with any feature-based method. In this work, we use the method proposed in [3].

Step 2: According to the 3D model, extract an self-occlusion map using Z-buffer.

Step 3: Evaluate the horizon points as described in definition 4.1.

Step 4: Compute gradient of evolution of the level-set function with respect to the 3D surface. Since level-set estimation methods are time and memory consuming, the gradient of the level-set is locally approximated. For a given point $\mathbf{q}$, the level-set is locally approximated by the tangent plane at $\mathbf{q}$ so that the gradient can be approximated as:

$$\psi(\varphi) \approx \nabla_\varphi(\mathbf{N}_\Gamma^\top(\mathbf{q}) \varphi(q)) \qquad (18)$$

Step 5: Update the surface function φ using gradient descent.

Step 6: Repeat step 2 to step 4 until a minimum is reached.

5. Experimental Results

The proposed method is tested on one synthetic and two real world scenes. The goal of these experiments is to show the influence of the proposed pixel-based method on the improvement of the self-occlusion map and on the reconstructed 3D surface. The proposed method does not depend on the type of the deformable surface and $\mathcal{E}_p$ can encode any physical prior. However, in feature-based methods only isometric surfaces has been defined in a variational framework such as:

$$
\mathcal{E}_p = \int_{\Omega} \|\mathsf{J}(\mathbf{q}_i)^{\top}\mathsf{J}(\mathbf{q}_i) - \mathbf{1}\|^2 d\mathbf{q}
$$
$$
+\alpha \int_{\Omega} \sum_{j=1}^{3} \sum_{k=1}^{2} \left\| \frac{\partial \varphi_j^i}{\partial q_{i_{(k)}}}(\mathbf{q}_i) \right\|^2 d\mathbf{q} \tag{19}
$$

Where the first term constraint the deformation to be isometric and the second weighted with the real number α is used to smooth up the reconstructed surface [3]. In our experiments, The surface function φ is initialized with the feature-based method described in [3]. This initial 3D reconstruction allows us to determine the initial visible and self-occluded sets. A first step optimization allows us to deform the surface to better fit the color discrepancy measure. The constraints (16) and (17) ensure that the 3D reconstruction consistent with the visible/self-occlusion constraints which link the surface and the vantage point.

5.1. Simulated Data

The Mandrill simulated dataset consists of 5 synthetic images of size 480×500 pixels. They are designed to evaluate the performance of our approach with different amounts of self-occlusion areas. A planar textured surface is deformed isometrically and projected to an image plane so that we obtain 5 different amounts of self-occlusions areas between 20% and 45% percents of the surface area as shown in figure 3. The correpondences are generated and perturbed with gaussian noise of zero mean and 2 pixels of std. The target image intensities are perturbed with a gaussian noise of zero mean and 5 of std (The intensity values being in the integer interval 0-255 for the RGB channels). It can be observed that the proposed pixel-based approach is very useful when the self-occluded area becomes bigger. Indeed, in figure 4 left we can see that more the self-occluded area is big and more the feature-based method gives poor results. While the proposed pixel-based approach gives results very close to the ground-truth self-occlusion map. The figures 4 middle and right evaluate quantitatively the performance of the proposed method regarding to the 3D reconstructed surface. It can be seen that the performance of the feature-based method depends highly on the number of the correspondences and their presence near the self-occlusion

boundary. For instance, this can be seen by comparing the performance between frame 1 and frame 5.

5.2. Real Data

We show the performance of the proposed approach for two datasets. Each dataset has a template image and a target image.

The 'Graffiti' dataset has a template image as shown in figure 5.a, of size 378×316 and a target image as shown in figure 5.b, of size 720×576. The image presents a deformed $A4$ paper sheet on which a graffiti is printed. Figures 5.c and 5.d show the result of the feature-based reconstruction where it can be seen that the reconstructed surface does not fit accurately to the deformed paper. The visible part in figure 5.c appears to be bigger than what is in the target image. Figures 5.e and 5.f exhibit the result obtained with the proposed pixel-based approach. It appears that in this case the visible part exhibited in figure 5.e fits more accurately the visible area on the surface in the target image. Indeed, it can be observed that the reconstructed surface in figure 5.f fits better the deformed paper.

The 'Toy-cloth' dataset results are shown in figure 1. In this dataset similar observations can be made as for our 'Graffiti' dataset.

6. Conclusion

We proposed a variational framework for pixel-based monocular template-based 3D reconstruction of a deformable surface. While state-of-the-art methods solve this problem using parametric formulations and feature-based methods, our approach is the first one that can perform simultaneous pixel-level image matching and 3D geometric reconstruction of the non-rigid surface.

We validated our approach on challenging real data sets showing phenomena such as surface self-occlusion on which previous methods tend to have poor accuracy.

References

[1] A. Bartoli and A. Zisserman. Direct estimation of non-rigid registrations. *BMVC*, 2004. 2

[2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. SIGGRAPH, 1999. 2

[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, 5, 6
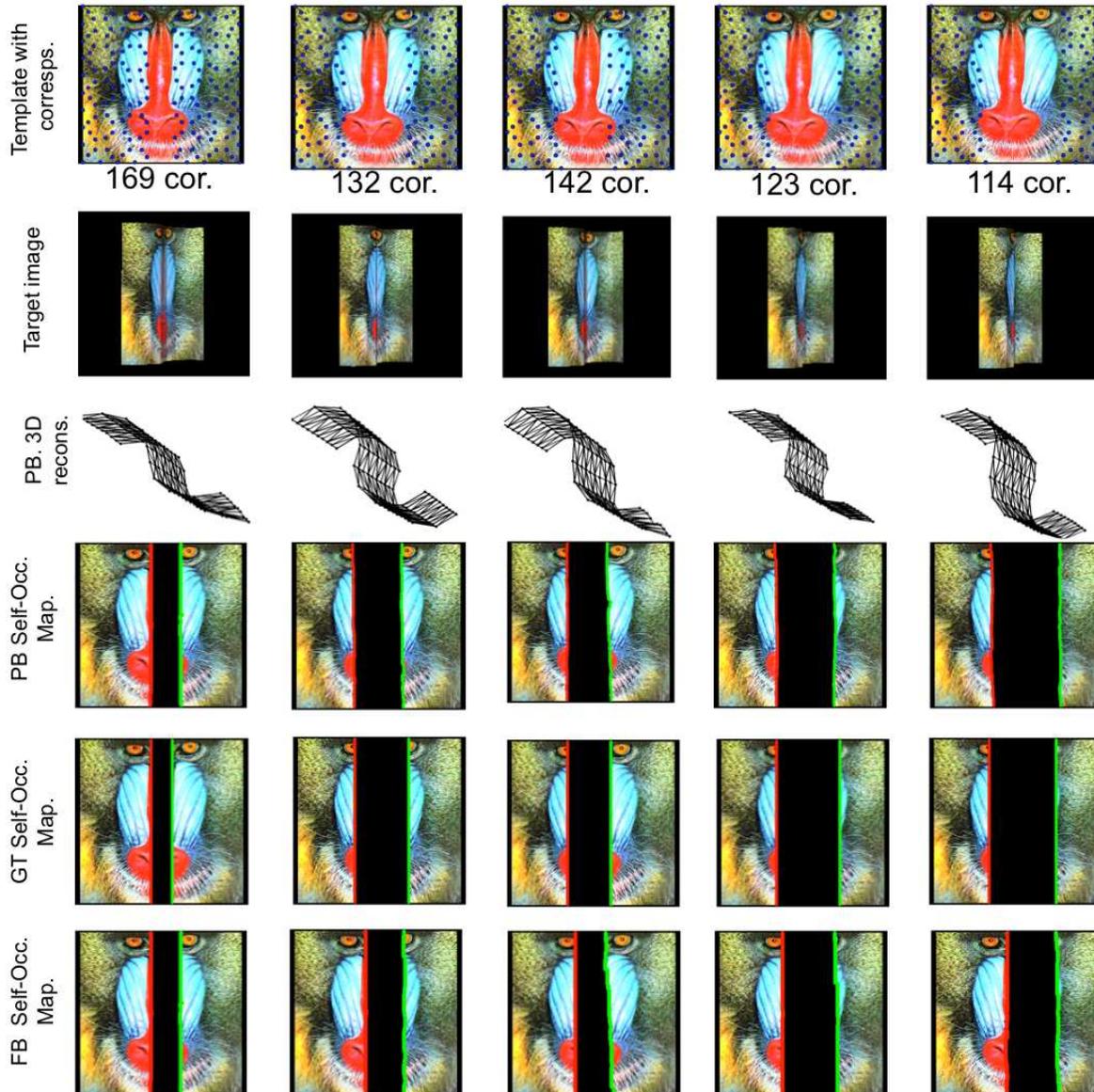
Figure 3. The "mandrill" simulated dataset. "GT" stands for ground-truth, "FB" stands for Feature-Based and "PB" stands for Pixel-Based.The blue points in the template images represent the correspondences used in the FB method. Their numbers are indicated below the template images. In the self-occlusion maps, the self-occluded area is in black. The horizon curve is in red and the terminator curve is in green.

[4] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7:336–344, 1998. 2

[5] P. Gargallo, E. Prados, and P. Sturm. Minimizing the reprojection error in surface reconstruction from images. *ICCV*, 2007. 2, 3

[6] V. Gay-Bellile, A. Bartoli, and P. Sayd. Direct estimation of non-rigid registrations with image-based self-occlusion reasoning. *PAMI*, 32(1):87–104, 2010. 2

[7] S. Ilic, M. Salzmann, and P. Fua. Implicit meshes for effective silhouette handling. *IJCV*, 72(2):159–178, 2007. 2

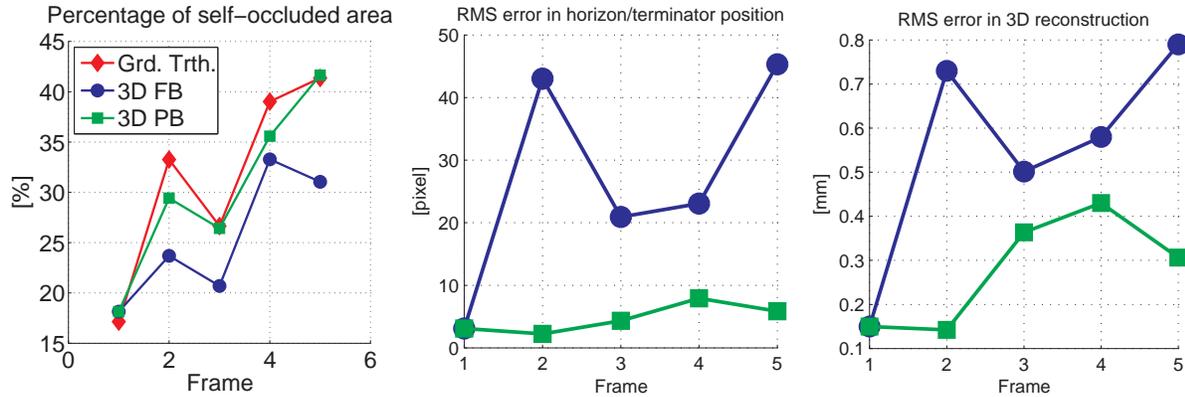[8] F. Moreno-Noguer, J.M. Porta, and P. Fua. Exploring

Figure 4. Left: Self-occlusion map expressed as the percentage of the self-occluded area above hole surface area. Middle: Horizon/Terminator position error after projection. Right: 3D reconstruction RMS error.
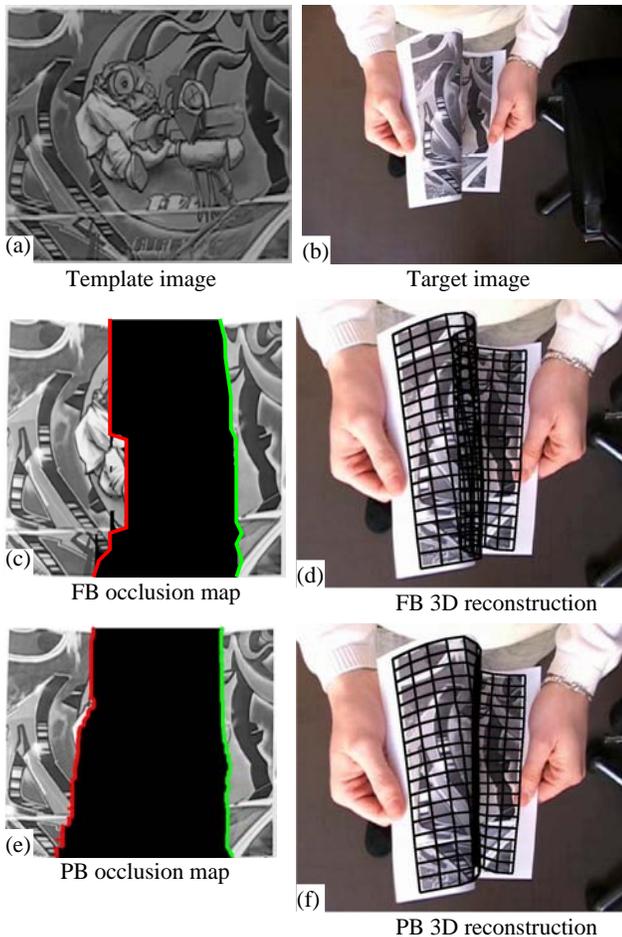

(a) Template image

(b) Target image

(c) FB occlusion map

(d) FB 3D reconstruction

(e) PB occlusion map

(f) PB 3D reconstruction

Figure 5. The "Graffitti" dataset. The horizon curve is in red and the terminator curve is in green. The self-occluded area is in black. "FB" stands for Feature-Based and "PB" stands for Pixel-Based.

ambiguities for monocular non-rigid shape estimation. *ECCV*, 2010. 2

[9] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *IJCV*, 95:124–137, 2011. 2

[10] J. Pilet, V. Lepetit, and P. Fua. Fast non-rigid surface detection, registration and realistic augmentation. *IJCV*, 76(2):109–122, 2008. 2

[11] D. Pizarro and A. Bartoli. Feature-based non-rigid surface detection with self-occlusion reasoning. *IJCV*, special issue: selected papers from 3DPVT'10, accepted April 2011, to appear. 2

[12] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. *CVPR*, 2009. 1, 2

[13] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *PAMI*, 2011. To appear. 2

[14] J. Solem and N. Overgaard. A geometric formulation of gradient descent for variational problems with moving surfaces. *Scale-Space*, 2005. 2, 3

[15] J. Taylor, A. D. Jepson, and K. Kutulakos. Non-rigid structure from locally-rigid motion. *CVPR*, 2010. 2

[16] Y. Tsai, L. Cheng, S. Osher, P. Burchard, and G. Sapiro. Visibility and its dynamics in a PDE based implicit framework. *J. Computational Physics*, 199:260–290, 2004. 3, 4

[17] H. Zhao, S. Osher, B. Merriman, and M. Kang. Implicit and nonparametric shape reconstruction from unorganized data using a variational level set method. *CVIU*, 80:295–314, 2000. 2