Linear Local Models for Monocular Reconstruction of Deformable Surfaces

Mathieu Salzmann, Pascal Fua

Abstract—Recovering the 3D shape of a nonrigid surface from a single viewpoint is known to be both ambiguous and challenging. Resolving the ambiguities typically requires prior knowledge about the most likely deformations that the surface may undergo. It often takes the form of a global deformation model that can be learned from training data. While effective, this approach suffers from the fact that a new model must be learned for each new surface, which means acquiring new training data and may be impractical.

In this paper, we replace the global models by linear local ones for surface patches, which can be assembled to represent arbitrary surface shapes as long as they are made of the same material. Not only do they eliminate the need to retrain the model for different surface shapes, they also let us formulate 3D shape reconstruction from correspondences as either an algebraic problem that can be solved in closed-form or a convex optimization problem whose solution can be found using standard numerical packages.

We present quantitative results on synthetic data, as well as qualitative ones on real images.

Index Terms—Deformable surfaces, Monocular shape recovery, Deformation models

1 Introduction

Being able to recover the 3D shape of deformable surfaces using a single camera would make it possible to field reconstruction systems that run on widely available hardware. However, because many different 3D shapes can have virtually the same projection, such monocular shape recovery is inherently ambiguous. The solutions that have been proposed over the years mainly fall into two classes: Those that involve physics-inspired models [32], [8], [19], [18], [22], [21], [35], [3] and those that learn global models from training data [9], [4], [7], [6], [1], [33], [17], [2], [15], [36], [39], [28]. The former solutions often entail designing complex objective functions and require hard-to-obtain knowledge about the precise material properties of the target surfaces. The latter require vast amounts of training data, which may not be available either, and only produce models for specific object shapes. As a consequence, one has to learn a specific deformation model for each individual object, even when all objects are made of the same material.

To overcome these limitations, we note that

- locally all parts of a physically homogeneous surface obey the same deformation rules;
- the local deformations are more constrained than those of the global surface and can be learned from fewer examples.

To take advantage of these facts, we represent the manifold of local surface deformations, and regularize the reconstruction of a global surface by encouraging its patches to conform to the local models. As shown in Fig. 1, this allows us to recover complex surface deformations for surfaces made of different materials from single *input* images when correspondences can be established with a *reference* image in which the surface shape is known.

In earlier work [29], we used nonlinear Gaussian Process Latent Variable Models to represent the space of local surface deformations. This has proved effective to recover the 3D deformations of relatively featureless surfaces from images from which only limited shape information can be extracted. This ability, however, came at a price: Using nonlinear deformation models results in highly non-convex objective functions, which requires good initialization. Furthermore, truly capturing the behavior of a material stills requires acquiring training data, which involves a painstaking motion capture process.

In this work, we advocate using simpler linear models instead to represent the local deformations in conjunction with inextensibility constraints. We show that, depending on whether the constraints are formulated as equalities or inequalities on distances between vertices of the mesh that represents the surface, reconstruction can be formulated either as a algebraic problem that can be solved in closed form or as a convex one whose solution can be found using standard numerical routines [5]. Either way, this relieves us from the need of an initialization and allows automatic reconstruction of sharply folding shapes such as those of Fig. 1 from *single images*. Furthermore, this entails no loss of accuracy with respect to the nonlinear models, especially when using inequality constraints as we first proposed in [25] rather than the equality constraints we introduced in [27]. Finally, if necessary, the linear models can be learned from synthetically generated data without even having to acquire motion capture data, which makes our approach practical even when such motion capture cannot be performed.

In short, we propose a generally applicable approach to recovering 3D shape from single images that is fully automated and can handle very complex deformations including sharp

- *M. Salzmann is with the Toyota Technological Institute at Chicago, IL, 60637, USA.*
- *P. Fua is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale, 1015 Lausanne, Switzerland.*

This work was supported in part by the Swiss National Science Foundation.

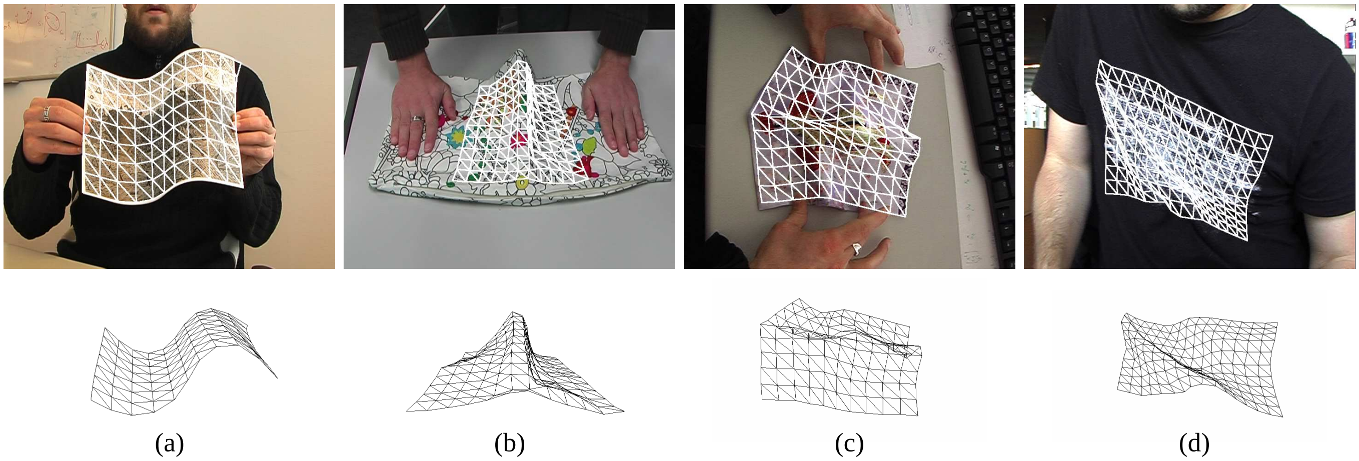


Fig. 1. Reconstruction of deformable surfaces undergoing complex deformations. Top Row: Reconstructed 3D mesh overlaid on the input image. Bottom Row: Side view of the same mesh.

folds and potentially featureless parts of the surface which we believe to be beyond the current state-of-the-art.

2 Related Work

3D reconstruction of nonrigid surfaces from single images is a severely under-constrained problem since many different shapes can produce very similar projections. Many methods have therefore been proposed over the years to favor the most likely shapes and disambiguate the problem.

The earliest approaches were inspired by physics and involved minimizing the sum of an internal energy representing the physical behavior of the surface and an external one derived from image data [32]. Many variations, such as balloons [8], deformable superquadrics [19] and thin-plates under tension [18], have since been proposed. Modal analysis has been applied to reduce the number of degrees of freedom of the problem by modeling the deformations as linear combinations of vibration modes [22], [21]. Since these formulations oversimplify reality, especially in the presence of large deformations, more accurate nonlinear models were proposed [35], [3]. However, to correctly reflect reality, these models need to be carefully hand-crafted, and give rise to highly nonlinear energy terms. In short, even though incorporating physical laws into the algorithms seems natural, the resulting methods suffer from two major drawbacks. First, one must specify material parameters that are typically unknown. Second, making them accurate in the presence of large deformations requires designing very complex objective functions that are often difficult to optimize.

Methods that learn global models from training data were introduced to overcome these limitations. As in modal analysis, surface deformations can be expressed as linear combinations of deformation modes. These modes, however, are obtained from training examples rather than from stiffness matrices and can therefore capture more of the true variability. For faces, Active Appearance Models [9] pioneered this approach in 2D and were quickly followed by 3D Morphable Models [4]. In previous work [28], we used a similar approach for general nonrigid surfaces and introduced a practical way of generating synthetic training data.

Nonrigid structure-from-motion methods also rely on learned linear models to constrain the relative motion of 3D points. Early approaches [7], [1] used known basis vectors, but the idea was expanded to simultaneously recover the shape and the modes from image sequences [6], [33], [39], [2], [15], [38]. However, since they rely on tracking points over long sequences, these methods often fail in practice. Only very recently has this problem been alleviated by using hierarchical priors [34], which assumes that the image measurements and 3D shapes come from a common probability distribution whose parameters are unknown. In any event, while learning deformation modes online is a very attractive idea, the resulting methods are only effective for relatively small deformations since using a large number of deformation modes makes the solution more ambiguous. Furthermore, whether learned offline or online, global models have the drawback of only being valid for a particular surface shape.

Recently, we proposed to replace the global deformation models by local ones that can be learned from smaller amounts of training data [29]. We represented the deformations of local patches of a surface with Gaussian Process Latent Variable Models (GPLVM) [13], and showed that a global deformation prior could be obtained by combining the local ones following a Product of Experts (PoE) [12] paradigm. This let us build models valid for any shape made of a particular material, and thus avoided the need to learn a new model for every new object shape. However, using a nonlinear representation of the local deformation yields non-convex objective functions. Therefore, to be effective, these models require good initialization and can only be used for tracking purposes.

Several methods have recently been proposed to recover the shape of inextensible surfaces without an explicit deformation model. Some are specifically designed for applicable surfaces, such as sheets of paper [11], [14], [23]. Others explicitly incorporate the fact that the distances between surface points must remain constant as constraints in the reconstruction process [27], [10], [24], [31]. This approach is very attractive because many materials do not perceptibly shrink or stretch as they deform. However, in our experience, additional regularization is still required when the surface is not textured

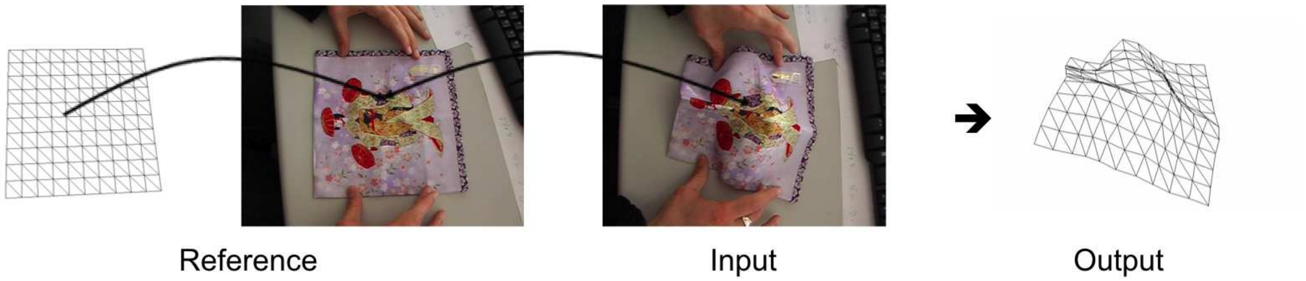


Fig. 2. Establishing 3D-to-2D correspondences. Given the reference mesh and image, we compute correspondences between 3D mesh locations given in barycentric coordinates and 2D feature points. From a new input image, we compute SIFT [16] matches with the reference image, which links the 3D surface points to 2D locations on the input image. The 3D shape is then obtained by deforming the mesh to make the 3D points best reproject on the input image.

enough. Furthermore, as will be discussed below, the constant distance assumption may be violated in the presence of sharp folds, which introduces inaccuracies.

3 Approach and Formulation

In this paper, we present a method that combines the strengths of inter-vertex distance constraints with those of local deformation models. It incorporates the following ingredients:

- **Shape from correspondences:** We show that reconstructing 3D shape from 3D-to-2D correspondences amounts to solving an ill-conditioned linear problem.
- **Linear local models:** To regularize the reconstruction and handle untextured surface parts, we introduce linear local models that can be learned either from motion-capture data or from easy-to-generate synthetic training data.
- **Inter-Vertex Distance Constraints:** Distance constraints are inherently non-linear and therefore not effectively enforced by the linear models. We therefore introduce them as non-linear constraints in our optimization scheme. We will show that this results in either an algebraic problem that can be solved in closed-form or a convex optimization problem, depending on whether the constraints are formulated as equalities or inequalities.

In the remainder of the paper, we discuss each one of these three ingredients in more detail. We then evaluate quantitatively the resulting algorithms.

To this end, we represent a surface as a triangulated mesh made of n_v vertices $\mathbf{v}_i = [x_i, y_i, z_i]^T$, $1 \leq i \leq n_v$ connected by n_e edges. Let $\mathbf{X} = [\mathbf{v}_1^T, \cdots, \mathbf{v}_{n_v}^T]^T$ be the vector of coordinates obtained by concatenating the $\mathbf{v}_i$.

We assume that we are given a set of n_c 3D-to-2D correspondences between the surface and an image. As depicted by Fig. 2, each correspondence relates a 3D point on the mesh, expressed in terms of its barycentric coordinates with respect to the facet to which it belongs, and a 2D feature in the image.

Additionally, we assume the camera to be calibrated and, therefore, the matrix of intrinsic parameters $\mathbf{A}$ to be known. To simplify our notations without loss of generality, we express the vertex coordinates in the camera referential. Note that, since we allow all the mesh vertices to move simultaneously, rigid surface motion is possible.

4 Shape from Correspondences

In this section, we formulate 3D surface reconstruction from 3D-to-2D correspondences as a linear problem. We then show that the resulting linear system is ill-conditioned and thus requires additional constraints.

4.1 Linear Formulation

Following [26], we first show that, given a set of 3D-to-2D correspondences, the vector of vertex coordinates $\mathbf{X}$ can be found as the solution of a linear system.

Let $\mathbf{p}$ be a 3D point belonging to facet f with barycentric coordinates $[b_1, b_2, b_3]$. Hence, we can write it as $\mathbf{p} = \sum_{i=1}^{3} b_i \mathbf{v}_{f,i}$, where $\{\mathbf{v}_{f,i}\}_{i=1,2,3}$ are the three vertices of facet f. The fact that $\mathbf{p}$ projects to the 2D image location (u, v) can now be expressed by the relation

$$\mathbf{A}\left(b_1\mathbf{v}_{f,1} + b_2\mathbf{v}_{f,2} + b_3\mathbf{v}_{f,3}\right) = k \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} , \qquad (1)$$

where k is a scalar accounting for depth. Since, from the last row of Eq. 1, k can be expressed in terms of the vertex coordinates, we have

$$\begin{bmatrix} b_1\mathbf{H} & b_2\mathbf{H} & b_3\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{f,1} \\ \mathbf{v}_{f,2} \\ \mathbf{v}_{f,3} \end{bmatrix} = \mathbf{0} , \qquad (2)$$

with

$$\mathbf{H} = \mathbf{A}_{\mathbf{2}\times\mathbf{3}} - \begin{bmatrix} u \\ v \end{bmatrix} \mathbf{A_3} , \qquad (3)$$

where $\mathbf{A}_{\mathbf{2}\times\mathbf{3}}$ contains the first two rows of $\mathbf{A}$, and $\mathbf{A_3}$ is the third one. n_c such correspondences between 3D surface points and 2D image locations therefore provide $2n_c$ linear constraints such as those of Eq. 2. They can be jointly expressed by the linear system

$$\mathbf{M}\mathbf{X} = \mathbf{0} , \qquad (4)$$

where $\mathbf{M}$ is a $2n_c \times 3n_v$ matrix obtained by concatenating the $\begin{bmatrix} b_1\mathbf{H} & b_2\mathbf{H} & b_3\mathbf{H} \end{bmatrix}$ matrices of Eq. 2.

Although solving the system of Eq. 4 yields a surface that reprojects correctly on the image, there is no guarantee that its 3D shape corresponds to reality. Indeed, not only is the rank of $\mathbf{M}$ not full due to the well-known global scale ambiguity, but,

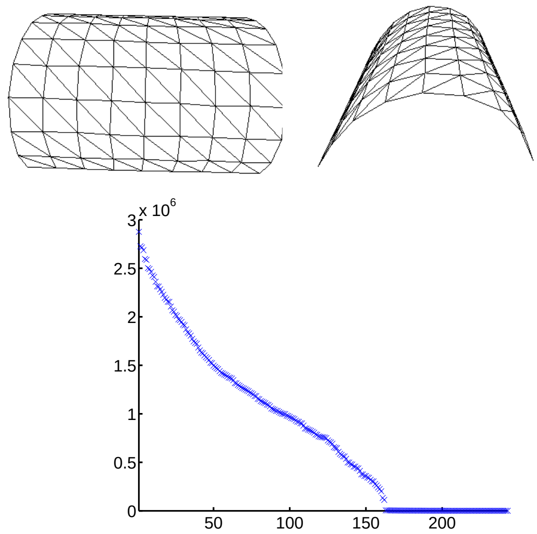


Fig. 3. Top row: Original and side views of a surface used to generate a synthetic sequence. The 3D shape was reconstructed by an optical motion capture system. Bottom row: Eigenvalues corresponding to the linear system of Eq. 4 written from correspondences randomly established for the mesh of the top left figure. The system was written in terms of 243 vertex coordinates. One third of the eigenvalues are close to zero.

for all practical purposes, it is even lower. More specifically, even where there are many correspondences, one third, i.e. n_v, of the eigenvalues of $\mathbf{M}^T\mathbf{M}$ are very close to zero, as illustrated by Fig. 3. In [26], we showed that this corresponds to one depth ambiguity per mesh vertex. As a result, even small amounts of noise produce large instabilities in the recovered shape.

This suggests that additional constraints have to be added to guarantee a unique and stable solution. In the following, we will show that using linear local deformation models in conjunction with inter-vertex distance constraints does the job and yields effective solutions.

5 Linear Local Models

In this section we introduce our surface deformation model and show that it lets us introduce a regularization term that greatly constrains the deformations the surface can undergo. However, this does not remove all ambiguities, which makes the length constraints of Section 6 necessary.

5.1 Learning Local Models

Representing the shape of a non-rigid surface as a linear combination of basis vectors is a well-known technique. Such a deformation basis can be obtained by modal analysis [22], [21], from training data [9], [4], [28], or directly from the images [39], [2], [15], [34], [38].

As shown in Fig. 4, we follow a similar idea, but, rather than introducing a single model for the whole surface, we subdivide the mesh into sets of overlapping patches and model the deformation of each one as a linear combination of modes. This lets us derive a deformation energy for each patch, and we take the overall mesh deformation energy to be the sum of those. In the appendix, we use motion capture data to provide empirical evidence that an energy formulated in this manner can be understood as the negative log of a shape prior.

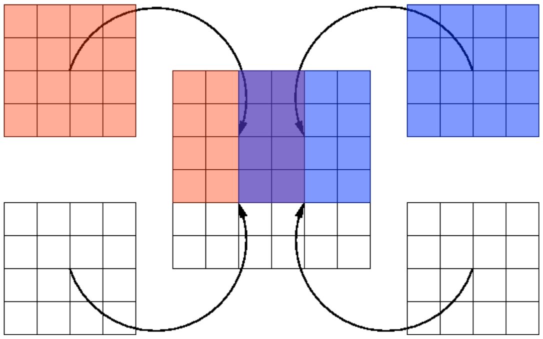


Fig. 4. Instead of modeling the whole surface, we subdivide the mesh into overlapping patches and model their deformations as linear combinations of modes. This lets us represent surfaces of arbitrary shape or topology by adequately assembling local patches.

Assuming that all parts of the surface follow similar deformation rules, the modes are the same for all patches and can be learned jointly, which minimizes the required amount of training data. Since patches can be assembled into arbitrarily-shaped global meshes, only one deformation model need be learned, irrespective of mesh shape and topology. Furthermore, local models also let us explicitly account for the fact that parts of the surface are much less textured than others and should therefore rely more strongly on the deformation model. This would not be possible with a global representation. Depending on parameter settings, it would either penalize complex deformations excessively, or allow the poorly textured regions to assume unlikely shapes.

Let $\mathbf{X}_i$ be the x-, y-, z-coordinates of an $n_l \times n_l$ square patch of the mesh. We model the variations of $\mathbf{X}_i$ as a linear combination of n_m modes, which we write in matrix form as

$$\mathbf{X}_i = \mathbf{X}_i^0 + \Lambda \mathbf{c}_i \ , \tag{5}$$

where $\mathbf{X}_i^0$ represents the coordinates of the patch in the reference image, Λ is the matrix whose columns are the modes, and $\mathbf{c}_i$ is the corresponding vector of mode weights. In practice, the columns of Λ contain the eigenvectors of the training data covariance matrix, and were computed by performing Principal Component Analysis on a set of deformed 5×5 meshes. As in [28], these meshes were obtained by simulating inextensible deformations. More specifically, we assigned random values uniformly sampled in the range $[-\pi/6, \pi/6]$ to a determining subset of the angles between the facets of the mesh. Some of the resulting modes are depicted in Fig. 5. Note that the same modes were used for *all* our experiments, independently of the material or shape of the surface of interest.

In [29], we introduced nonlinear local models. While they offer a more accurate representation of the space of possible deformations, which is known to be nonlinear, they suffer from two drawbacks. First, they yield a highly non-convex shape likelihood function, which only makes them practical

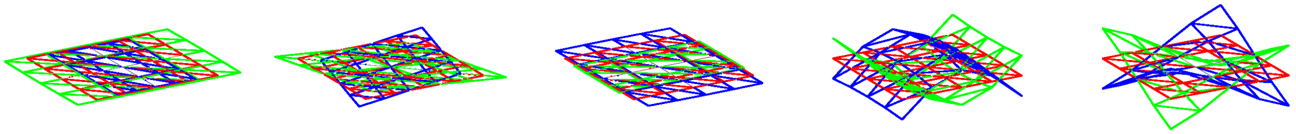


Fig. 5. Visual interpretation of the local deformation modes. We show the effect of adding (blue) or subtracting (green) some of the modes to the mean shape (red). Note that, despite the fact that all the training examples were inextensible deformations of a mesh, PCA yields extension modes.

for tracking purposes. Second, to accurately capture the space of feasible deformations of a particular material, they need training examples acquired from a real object, which involves a painstaking process. Our linear local models have the advantage that they can be learned from synthetic training data, that can easily be generated. Furthermore, as long as sufficiently many modes are kept, they define an hyper-ellipsoid that encompasses the true nonlinear deformation space. Therefore, they can model arbitrarily complex shapes. In practice, to remain as general as possible, we keep *all* the modes and enforce deformations to remain plausible by regularizing their coefficients according to their importance, as described below.

5.2 Local Models for Shape Recovery

When using a linear model for shape recovery, the usual approach is to replace the original unknowns by the modes weights. However, since we model the global surface with overlapping local patches, doing so would not constrain the shapes predicted by the weights associated to two such patches to be consistent. Fortunately, since the deformation modes are orthonormal, the coefficients $\mathbf{c}_i$ of Eq. 5 can be directly computed from $\mathbf{X}_i$ as

$$\mathbf{c}_i = \Lambda^T \left(\mathbf{X}_i - \mathbf{X}_i^0\right) \ . \tag{6}$$

We therefore use the vector of surface coordinates $\mathbf{X}$ introduced in Section 4.1. To enforce the individual surface patches to conform to our linear local model, we use all the modes and introduce the penalty term

$$\left\|\Sigma^{-1/2}\mathbf{c}_i\right\| = \left\|\Sigma^{-1/2}\Lambda^T\left(\mathbf{X}_i - \mathbf{X}_i^0\right)\right\| \ , \tag{7}$$

where Σ is a diagonal matrix that contains the eigenvalues associated to the eigenvectors in Λ. It measures how far the $\mathbf{c}_i$, and therefore the $\mathbf{X}_i$, are from the training data. We then write the global regularization term as the solution to the optimization problem

$$\underset{\mathbf{X}}{\text{minimize}} \left\|\mathbf{W}_l \mathbf{L}\left(\mathbf{X}-\mathbf{X}^0\right)\right\|^2 \ , \tag{8}$$

where $\mathbf{L}$ is an $n_p n_l^2 \times n_v$ matrix which concatenates n_p copies of $\Sigma^{-1/2}\Lambda^T$ spread over the global mesh $\mathbf{X}$ according to the vertices of the n_p patches $\mathbf{X}_i$, and $\mathbf{X}^0$ is the reference shape of the global mesh. $\mathbf{W}_l$ is a diagonal matrix containing n_p individual values w_l^i designed to account for the fact that poorly-textured patches should rely more strongly on the model than well-textured ones. In other words, w_l^i should be inversely proportional to the number of correspondences in patch i. We take it to be

$$w_l^i = \exp\left(-\frac{n_{in}^i}{\text{median}(n_{in}^k > 0\ ,\ 1 \le k \le n_p)}\right) \ , \tag{9}$$

where n_{in}^j is the number of inlier matches in patch j. Note that the formulation of the shape regularization of Eq. 8 spares us the need to explicitly introduce additional latent variables as was the case for the nonlinear local models [29].

To prevent us from obtaining the trivial solution $\mathbf{X} = \mathbf{X}^0$ to the problem of Eq. 8, we solve it in conjunction with the projection equations of Eq. 4. This lets us express the shape reconstruction problem as the solution of

$$\underset{\mathbf{X}}{\text{minimize}} \left\|\mathbf{M}\mathbf{X}\right\|^2 + \left\|\mathbf{W}_l \mathbf{L}\left(\mathbf{X}-\mathbf{X}^0\right)\right\|^2 \ . \tag{10}$$

Since, within the L_2-norms, both terms are linear in $\mathbf{X}$, this is equivalent to solving in the least-squares sense the linear system

$$\mathbf{S}\begin{bmatrix}\mathbf{X}\\ 1\end{bmatrix} = \mathbf{0}\ , \tag{11}$$

where

$$\mathbf{S} = \begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{W}_l\mathbf{L} & -\mathbf{W}_l\mathbf{L}\mathbf{X}^0\end{bmatrix} \ . \tag{12}$$

In Fig. 6, we plot the eigenvalues $\mathbf{S}^T\mathbf{S}$ for the mesh of Fig. 3. As we can see, much fewer eigenvalues are close to zeros than before. This suggests that our linear local models truly improve the conditioning of our problem. However, some eigenvalues remain small, which implies that some ambiguities are still unresolved. This, for example, is the case of the global scale ambiguity that can be modeled by the extension modes depicted in Fig. 5. Therefore, additional constraints need to be introduced to fully disambiguate the problem.

6 Nonlinear Constraints

In this section, we introduce the additional nonlinear constraints that, in conjunction with the linear local models of the previous section, make shape recovery from 3D-to-2D correspondences well-posed. We first introduce inextensibility constraints, and show that they yield a closed-form solution of the reconstruction problem. Then, because these constraints may be violated in the presence of sharp folds, we replace them by distance inequalities, which results in a convex formulation.

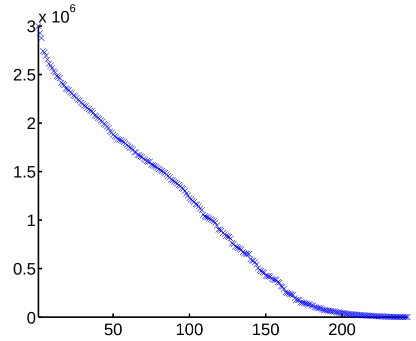


Fig. 6. Eigenvalues corresponding to the linear system of Eq. 11 for the mesh of Fig. 3. Note that fewer eigenvalues are close to zero than when relying on texture only. However, some remain small, which suggests that the linear local models do not fully disambiguate the problem.

6.1 Distance Equality Constraints

Several recent approaches [24], [10], [27] rely on the fact that many deformable surfaces, such as clothes or paper, are nearly inextensible. In our case, this means enforcing constraints expressed as

$$\|\mathbf{v}_j - \mathbf{v}_k\|^2 = l_{j,k}^2 \ , \ \forall (j,k) \in \mathcal{E} \ , \tag{13}$$

where $\mathcal{E}$ represents the set of n_e edges of the mesh, and $l_{j,k}$ is the length of the edge joining vertex j and vertex k in the reference configuration. A typical way to solve such quadratic constraints in closed-form is to linearize the system, which involves introducing new unknowns for the quadratic terms. In our case, this would yield $3n_v(3n_v+1)/2$ unknowns, which, for meshes of reasonable size, would quickly become intractable. Instead, we propose to describe the solutions of Eq. 11 with a reduced number of unknowns, which lets us effectively enforce inextensibility constraints.

Following the idea introduced in [20], we write the solution of the linear system of Eq. 11 as a weighted sum of the eigenvectors $\mathbf{s}_i \ , 1 \leq i \leq n_s$ of $\mathbf{S}^T\mathbf{S}$, which are associated with the n_s smallest eigenvalues. Therefore we write

$$\begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \sum_{i=1}^{n_s} \beta_i \mathbf{s}_i \ , \tag{14}$$

since any such linear combination of $\mathbf{s}_i$ is in the kernel of $\mathbf{S}^T\mathbf{S}$ and produces a mesh that simultaneously projects correctly on the image and conforms to the linear local models. Our problem now becomes one of finding appropriate values for the β_i, which are the new unknowns.

We are now in a position to exploit the inextensibility of the surface by choosing the β_i so that edge lengths are preserved. Such β_i can be expressed as the solution of a set of quadratic equations of the form

$$\|\sum_{i=1}^{n_s} \beta_i \mathbf{s}_i^j - \sum_{i=1}^{n_s} \beta_i \mathbf{s}_i^k\|^2 = l_{j,k}^2 \ , \tag{15}$$

where $\mathbf{s}_i^j$ is the 3×1 sub-vector of $\mathbf{s}_i$ corresponding to the coordinates of vertex $\mathbf{v}_j$. In addition to these quadratic constraints, we need to express the fact that the last elements of the products $\beta_i \mathbf{s}_i$ must sum up to one. This yields the linear equation

$$\sum_{i=1}^{n_s} \beta_i \mathbf{s}_i^{3n_v+1} = 1 \ , \tag{16}$$

which we solve together with the quadratic edge constraints.

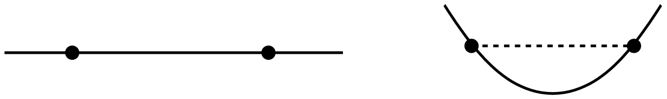


Fig. 7. Schematic representation of why inextensibility constraints are ill-suited for sharp folds. Left: Two points of the discrete representation of a continuous surface in its rest configuration. Right: When deformed, while the geodesic distance between the two points is preserved, the Euclidean one decreases. This suggests that distance inequality constraints should be used rather than equalities.

Since $n_s \ll 3n_v$, linearization becomes a viable option to solve our quadratic equations. To this end, we consider the quadratic terms as additional variables, and define the new $(n_s(n_s+3)/2)$-dimensional vector of unknows as $\mathbf{b} = [\mathbf{b_l}^T, \mathbf{b_q}^T]^T$, such that

$$\begin{aligned} \mathbf{b_l} &= [\beta_1, \cdots, \beta_{n_s}]^T \ , \text{ and} \\ \mathbf{b_q} &= [\beta_1\beta_1, \cdots, \beta_1\beta_{n_s}, \beta_2\beta_2, \cdots, \beta_2\beta_{n_s}, \cdots, \beta_{n_s}\beta_{n_s}]^T \ . \end{aligned}$$

Finding a shape that satisfies the constraints described above can now be expressed as solving the optimization problem

$$\underset{\mathbf{b}}{\text{minimize}} \ \|\mathbf{D}\mathbf{b_q} - \mathbf{d}\|^2 + w_s \left(\mathbf{s}^{3n_v+1}\mathbf{b_l} - 1\right)^2 \ , \tag{17}$$

where $\mathbf{D}$ is an $n_e \times n_s(n_s+1)/2$ matrix built from the known $\mathbf{s}_i$, $\mathbf{d}$ is the $n_e \times 1$ vector of edge lengths in the reference configuration, and $\mathbf{s}^{3n_v+1}$ is the row vector containing the last element of each s_i. w_s is a weight that sets the influence of the constraint of Eq. 16, and was always set to 1e6. Note that, with our new unknowns, this problem is equivalent to solving a linear system in the least-squares sense, which can be done in closed-form.

However, solving the problem of Eq. 17 directly would yield a meaningless solution since nothing links the linear terms with the quadratic ones. To overcome this problem, we multiply the linear equation of Eq. 16 by the individual β_j, which yields n_s new equations of the form

$$\sum_{i=1}^{n_s} \beta_j \beta_i \mathbf{s}_i^{3n_v+1} = \beta_j \ . \tag{18}$$

Adding these equations to Eq. 17 provides the missing link between linear and quadratic terms. Note that this does not truly guarantee consistency between the linear and quadratic terms, but, in practice, it proved sufficient to yield meaningful reconstructions. We therefore solve the optimization problem

$$\underset{\mathbf{b}}{\text{minimize}} \ \|\mathbf{D}\mathbf{b_q} - \mathbf{d}\|^2 + w_s \left(\left(\mathbf{s}^{3n_v+1}\mathbf{b_l} - 1\right)^2 + \|\mathbf{D_{lq}}\mathbf{b}\|^2\right) \ , \tag{19}$$

where $\mathbf{D_{lq}}$ is an $n_s \times n_s(n_s+3)/2$ matrix. Note that this problem can still be solved in closed-form. Given its solution,

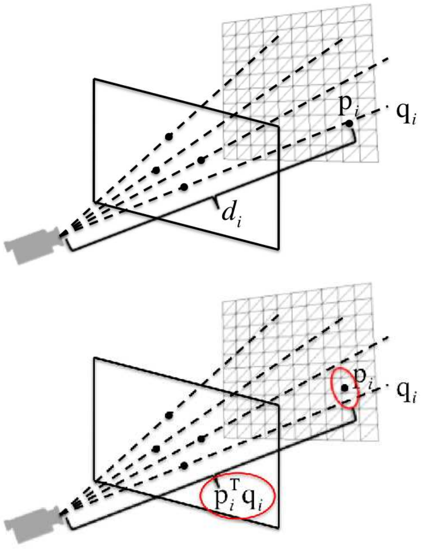


Fig. 8. With a perspective camera model, lines-of-sight are not parallel. Therefore, maximizing the area of a mesh can be achieved by pushing it away from the camera. **Top:** In the absence of noise this can be done by maximizing the depth of the point along the line-of-sight. **Bottom:** With noise, we replace the depth d_i by the projection of the point on the line-of-sight.

we can compute the shape of the deforming surface from Eq. 14 with the linear terms of vector $\mathbf{b}$. Selecting the correct number n_s of eigenvectors to take into account is done by testing for all values smaller than a predefined threshold, and by picking the one that gives the smallest mean edge length variation. In practice the maximum value for n_s was set to 20.

6.2 Distance Inequality Constraints

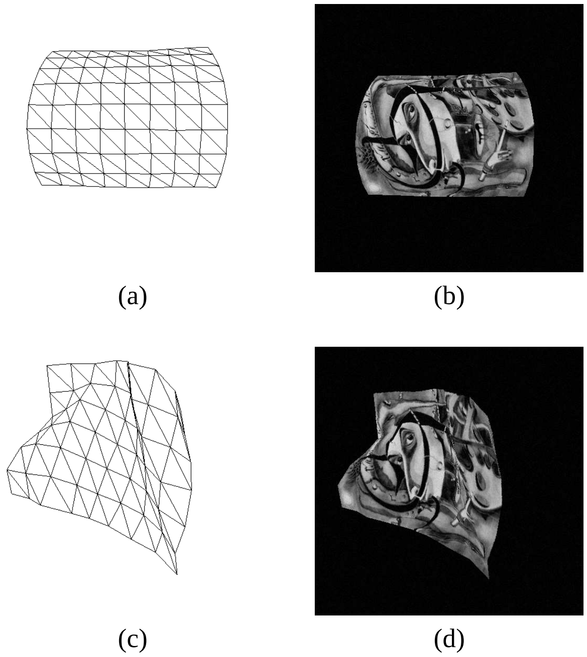
As we will show in the results section, the inextensibility constraints yield good reconstruction of smoothly deforming surfaces. However, as illustrated by Fig. 7, such constraints are violated when folds appear between mesh vertices, because the Euclidean distance between points on the surface may decrease. It is therefore truer to reality to replace the inextensibility constraints by constraints that allow vertices to come closer to each other, but not to move further apart than their geodesic distance [25]. For all pairs of neighboring vertices $\mathbf{v}_j$ and $\mathbf{v}_k$, we therefore replace the constraints of Eq. 13 by inequality constraints written as

$$\|\mathbf{v}_k - \mathbf{v}_j\| \leq l_{j,k} \ . \tag{20}$$

Note that, contrary to inextensibility constraints, these distance inequalities are convex [5]. As a consequence, there is no need to linearize them, and we could directly solve the problem

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \ \|\mathbf{M}\mathbf{X}\| + \left\|\mathbf{W}_l\mathbf{L}\left(\mathbf{X}-\mathbf{X}^0\right)\right\| \\ &\text{subject to} \ \ \|\mathbf{v}_k - \mathbf{v}_j\| \leq l_{j,k} \ , \ \forall (j,k) \in \mathcal{E} \ . \end{aligned} \tag{21}$$

This could be done using available convex optimization packages [30] by introducing a slack variable to minimize the norm [5].


(a) (b)

(c) (d)

Fig. 9. Synthetic data acquired with a motion capture system. (a,b) Mesh and corresponding textured image of a smoothly deforming piece of cardboard. (c,d) Similar images for a piece of cloth with sharper folds.

However, while our inequalities prevent the mesh from expanding, they still allow it to shrink to a single point. This could be remedied by maximizing the mesh area under our constraints. However, this would yield a non-convex problem. Instead, we exploit the fact that, in the perspective camera model, the lines-of-sight are not parallel, as depicted by the top drawing of Fig. 8. Thus the largest distance between two points is reached when the surface is furthest away from the camera. Therefore, a nontrivial reconstruction can be obtained by maximizing the depth d_i of each point along its line-of-sight $\mathbf{q}_i$. While, with noise-free correspondences, 3D surface points are completely defined by their position along the lines-of-sight, they should be allowed to move away from them in the presence of noise, as depicted by the bottom of Fig. 8. Therefore, rather than maximizing d_i, we consider the projections of $\mathbf{p}_i$ on its line-of-sight $\mathbf{q}_i$, which can be computed as

$$\mathbf{p}_i^T\mathbf{q}_i = \mathbf{X}^T\mathbf{B}_i^T\mathbf{q}_i \ , \tag{22}$$

where $\mathbf{B}_i$ is the $3 \times 3n_v$ matrix containing the barycentric coordinates of point i placed to correctly match the vertices of the facet to which the point belongs.

We can then add the maximization of the terms of Eq. 22 to the optimization problem of Eq. 21, which yields the new convex problem

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \ \|\mathbf{M}\mathbf{X}\| + \left\|\mathbf{W}_l\mathbf{L}\left(\mathbf{X}-\mathbf{X}^0\right)\right\| - w_d\sum_{i=1}^{n_{in}}\mathbf{X}^T\mathbf{B}_i^T\mathbf{s}_i \\ &\text{subject to} \ \ \|\mathbf{v}_k - \mathbf{v}_j\| \leq l_{j,k} \ , \ \forall (j,k) \in \mathcal{E} \ , \end{aligned} \tag{23}$$

where w_d is a weight that controls the relative influence of depth maximization and image error minimization. In practice,

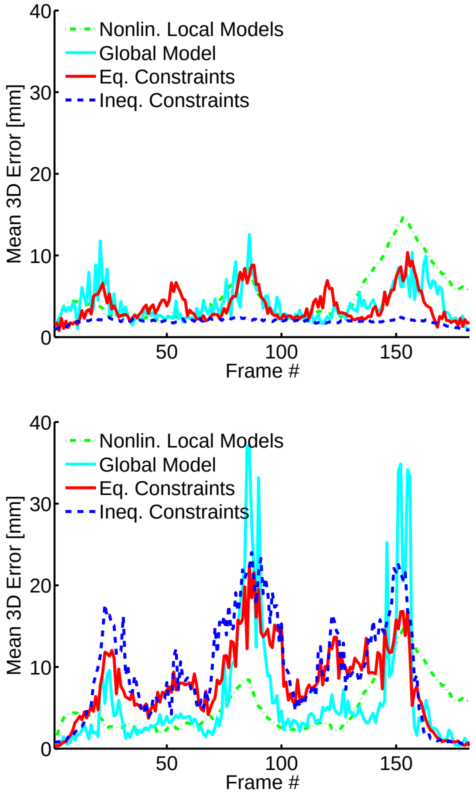


Fig. 10. Reconstruction error for the cardboard sequence. Mean vertex-to-vertex distance to ground-truth meshes from synthetic correspondences (top) and SIFT correspondences (bottom). We compare our results with those of the methods in [27] (cyan) and [29] (green). Results obtained with equality constraints are shown in red and with inequalities in blue.

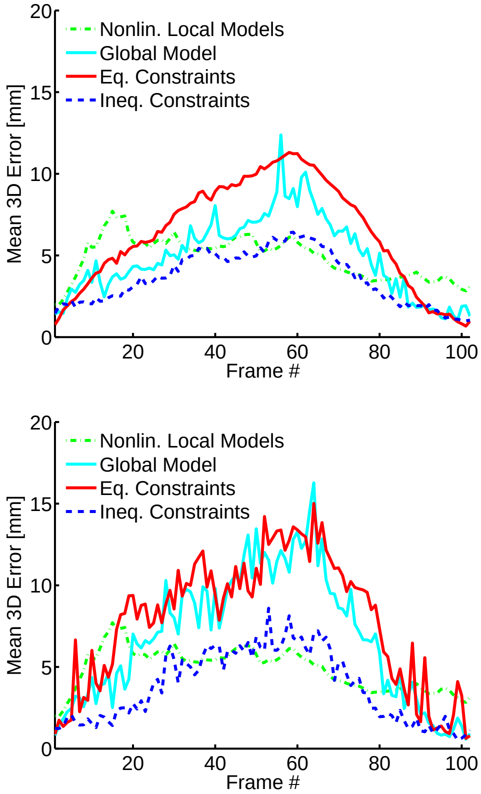


Fig. 11. Similar plots as in 10 for the deformations of a piece of cloth.

we set w_d to 2/3 because computing depths involves $3n_{in}$ values against $2n_{in}$ projection equations. Since we simply added linear terms to the previous objective function, this optimization problem remains convex.

7 Experimental Results

We now present results obtained on synthetic and real data by using our linear local models with either the inextensibility constraints of Section 6.1 or the distance inequalities of Section 6.2. Note that the meshes we used to produce these results all have different dimensions. Nevertheless, thanks to our local models, we only had to compute the deformation modes once for 5x5 pacthes and then to combine them appropriately for the different meshes.

7.1 Synthetic Data

We applied our two approaches to synthetic data to quantitatively evaluate their performance. Furthermore, we compare them against our closed-form solution relying on a global deformation model and inextensibility constraints [27], and against nonlinear local deformation models [29]. Note that the latter method relies on template matching instead of correspondences and tracks the deformation from frame to frame due to the non-convexity of its objective function.

To make our experiments as realistic as possible, we obtained 3D meshes, such as those of Fig. 9(a,c), by deforming a sheet of cardboard and a more flexible piece of cloth in front of an optical motion capture system. We then created correspondences in two different manners. We first created completely synthetic correspondences by randomly sampling the barycentric coordinates of the mesh facets, projecting them with a known camera, and adding zero-mean Gaussian noise with variance 2 to the image locations. To simulate real data even more accurately, we textured the meshes and generated images, such as the ones of Fig. 9(b,d), with uniform intensity noise in the range $[-10, 10]$. We then obtained correspondences by matching SIFT [16] features between a reference image and the input images. To cope with the outliers resulting from this procedure, we implemented an iterated reweighting procedure that decreases a radius inside which correspondences are considered as inliers. In practice, we initialized this radius to 50 pixels and divided it by 2 at each iteration. We then weighted each valid line of the matrix $\mathbf{M}$ of Eq. 4 by a weight

$$w_i = \exp\left(-\frac{e_i}{\text{median}(e_j\ ,\ 1 \leq j \leq n_{in})}\right)\ , \tag{24}$$

where e_i is the reprojection error of correspondence i, and n_{in} is the number of inliers. The same procedure was used with the synthetic outliers described below and with real images discussed in Section 7.2.

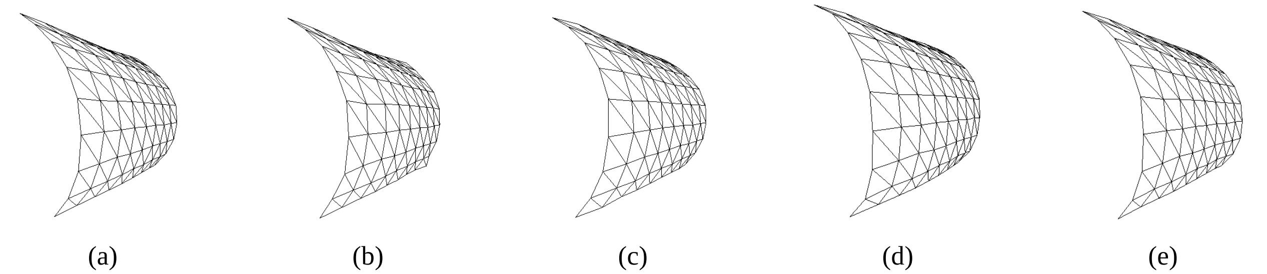


Fig. 12. Visual comparison of the recovered meshes for the deformation of Fig. 9(a). (a) Ground truth. Mesh recovered with (b) non-linear local models, (c) global model with equality constraints, (d) local models with equality constraints, (e) local models with inequality constraints. Beacause the deformation is fairly smooth, all recovered shapes are fairly similar.

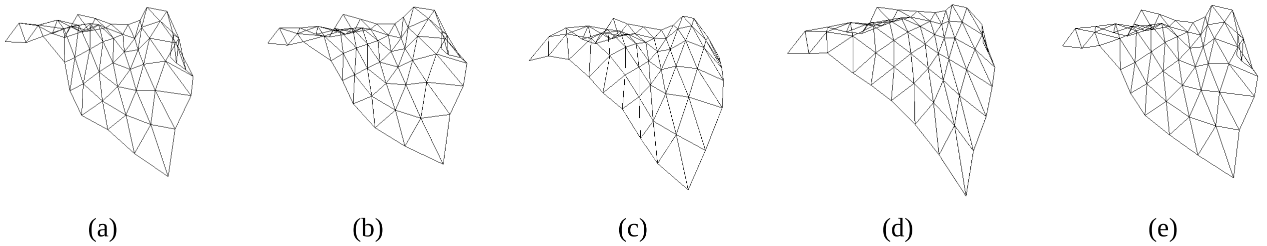


Fig. 13. Visual comparison of the recovered meshes for the deformation of Fig. 9(c). (a) Ground truth. Mesh recovered with (b) non-linear local models, (c) global model with equality constraints, (d) local models with equality constraints, (e) local models with inequality constraints. Because the folds are sharp, using equality constraints tends to oversmooth whereas inequalities or nonlinear models yields better results.

In Figs. 10 and 11, we compare the results of the four different techniques on the sheet of cardboard and the piece of cloth, respectively. We plot the mean vertex-to-vertex distance between the reconstructed mesh and the ground-truth one. On the top plot of each figure, we show the results obtained with synthetic matches, and on the bottom one, the errors obtained with SIFT matches. In Figs. 12 and 13, we visually compare the results of all approaches for the frames in which the deformation is largest, i.e. frames 100 and 60, respectively. From these curves, we can observe that using inequality constraints gives better results, especially for the piece of cloth. This was to be expected since sharp folds are better modeled by inequalities. Furthermore, we can observe that local and global models used in conjunction with equality constraints perform similarly. While this might seem disappointing, local models still have the advantage of being more general than the global ones in the sense that they let us model arbitrary shapes. Finally, while nonlinear local models perform well, they involve tracking the surface throughout the sequence, which can result in drift, as can be observed at the end of the cardboard sequence. Additionally, they are much more computationally expensive than the closed-form or convex optimization methods.

To test the robustness of our approaches to the lack of texture, we used the synthetic correspondences, and removed randomly selected subsets of them. In Fig. 14, we plot the average reconstruction error over the sequences as a function of the percentage of removed correspondences. As shown by the plots, accuracy does not decrease significantly until most correspondences are gone. Finally, we tested the robustness of our approach to outliers by assigning random image locations to a given percentage of the synthetic correspondences. In Fig. 15, we plot the mean reconstruction error over the sequences as a function of the outlier rate. As we can see, both methods are robust to up to 50% outliers. However, the distance equality constraints are more stable for higher outlier rates.

In Figs. 16 and 17, we show the limitations of our approach when there is little texture concentrated in a single area of the surface, which almost amounts to a worst-case scenario. To this end, we textured the same cardboard and cloth surfaces as before to create images such as the ones of Fig. 17(a,d), and computed sift correspondences from them. Fig. 16 depicts the reconstruction errors for the different frames of the sequences. Note that the values are significantly higher than those of Figs. 10 and 11. In Fig. 17(b,c,e,f), we plot the recovered 3D shapes for the same frames as in Fig. 12 and 13 to quantitatively evaluate these results. Note that the reconstructed surfaces are much flatter than before. This was to be expected since we only have shape information for the textured part, and suggests that additional image cues, such as edges or shading, should be used.

7.2 Real Images

We tested our approach on real images taken with a 3-CCD DV camera. In each one of the following figures, we show the mesh recovered overlaid on the input image and the same mesh seen from a different viewpoint. Note that, even though our results were obtained from video sequences, nothing links the shape recovered in the consecutive frames. We first used the

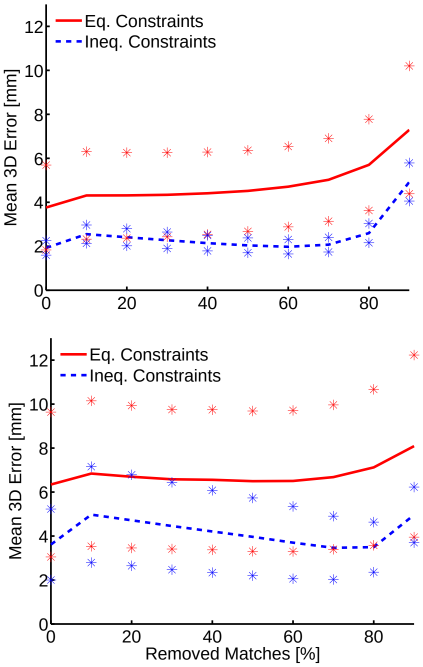


Fig. 14. To evaluate the influence of the lack of texture on our methods, we removed randomly selected subsets of correspondences. We plot the mean reconstruction error over the whole sequence as a function of the percentage of removed matches for the cardboard data (top) and the cloth sequence (bottom). The stars indicate the standard deviation of the error.

equality constraints to recover the deformations of smoothly deforming objects such as the sheets of paper of Fig. 18. In Fig. 19, we show that, if the mesh is fine enough, the equality constraints can still reconstruct folds. However, if the folds on the surface do not correspond to mesh edges as in the case in Fig. 20, these constraints are not appropriate anymore. As can be observed in the bottom row of the figure, the folds cannot be modeled correctly, and the recovered shapes are too smooth. This is not the case anymore with distance inequalities, as shown in the second row. Fig. 21 depicts results obtained with our distance inequality constraints on two other flexible surfaces. Finally, we applied our method to recover the shape of the non-rectangular surface depicted by Fig. 22. In this case, the correspondences were obtained by tracking markers on the sail. In Fig. 22(g), we show how we covered the entire sail with local models. Note that the additional vertices required by our local models have no negative influence on the recovered shapes since they do not contain any correspondences.

8 Conclusion

In this paper, we have presented linear local deformation models for 3D shape reconstruction from monocular images. We have shown that these models have the advantage of being more general than global ones, and of being easier to deploy than nonlinear local models. Furthermore, we have shown that, when used in conjunction with distance constraints, they yield accurate solutions to the shape recovery problem. In

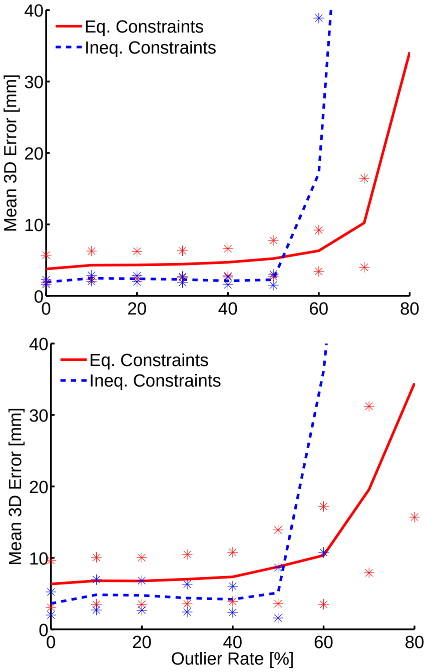


Fig. 15. We evaluated the robustness of our approaches to outliers by setting random values to the image locations of some correspondences. We plot the mean reconstruction error over the whole sequence as a function of the outlier rate for the cardboard data (top) and the cloth sequence (bottom). The stars indicate the standard deviation of the error.

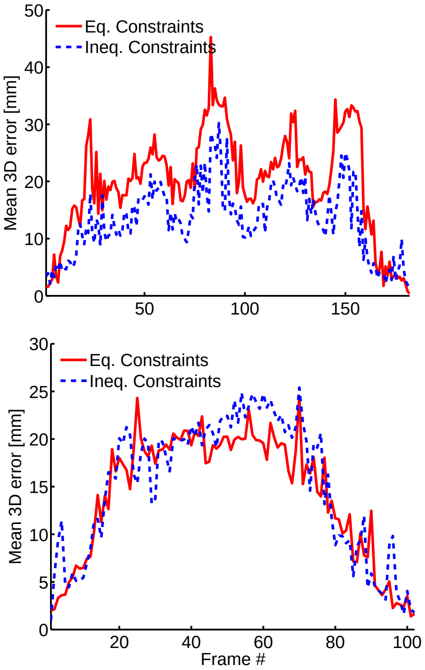


Fig. 16. Reconstruction errors from SIFT correspondences on the poorly textured surfaces of Fig. 17(a,d) for a piece of cardboard (left) and for a piece of cloth (right) Note that these errors are significantly larger than those of Figs. 10 and 11.

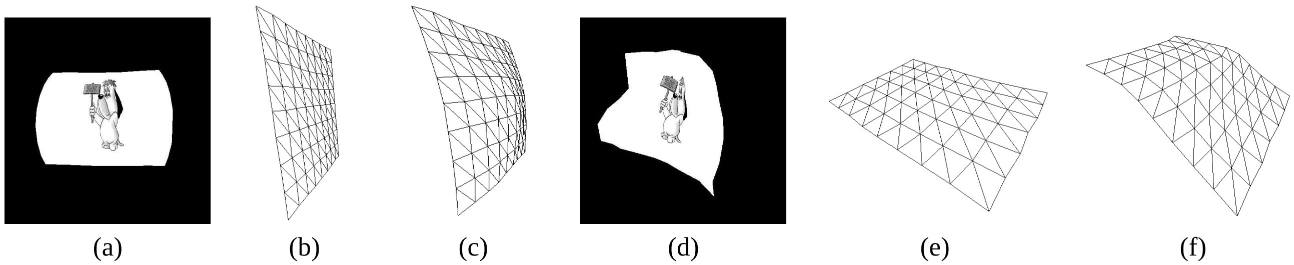


Fig. 17. Recovering the shape of poorly textured surfaces (a,d). (b,e) 3D reconstruction using equality constraints. (c,f) 3D reconstruction using inequality constraints. Since we only exploit shape information in the center of the image, the recovered surfaces are far too smooth.

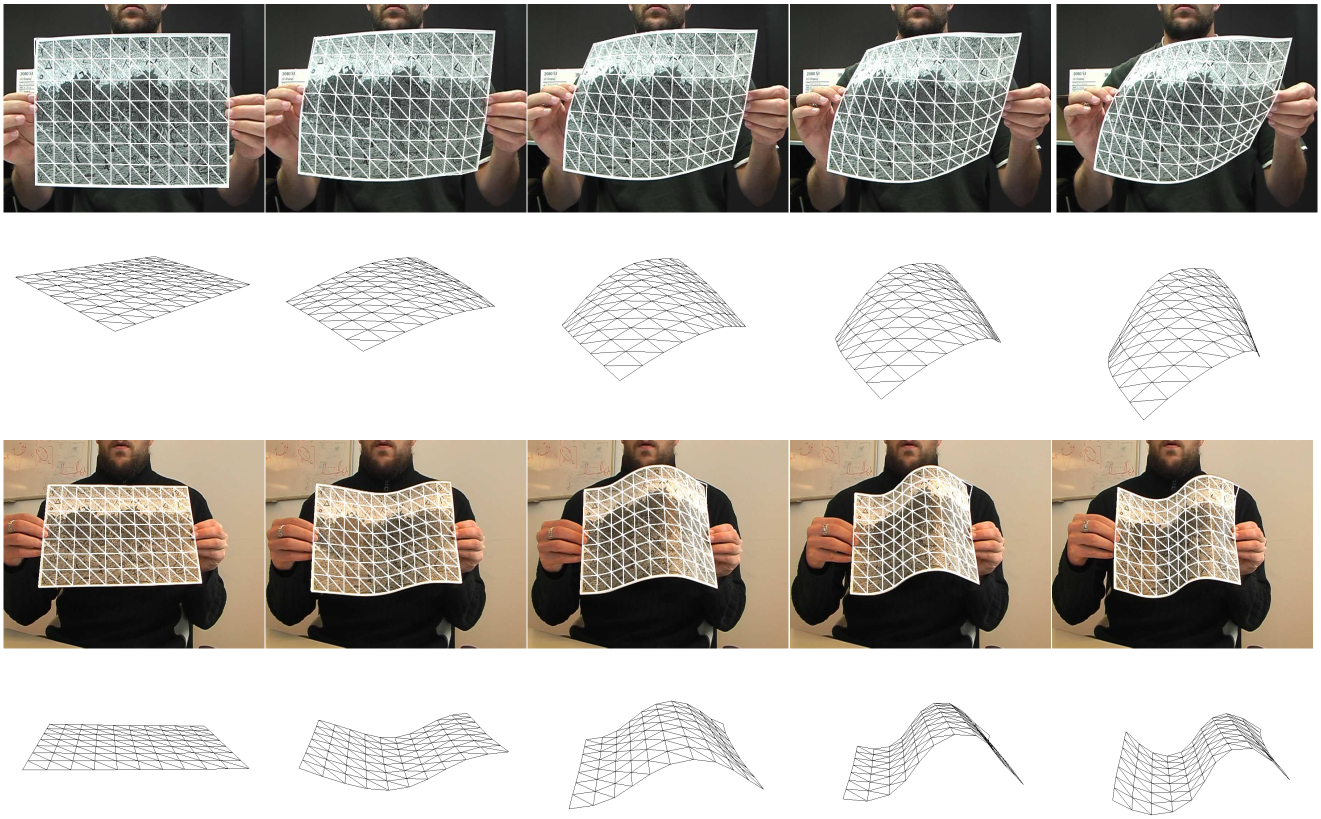


Fig. 18. Recovering the shape of a piece of paper. First and third rows: Mesh recovered using equality constraints overlaid on the input image. Second and fourth rows: Side view of that mesh.

particular, we have introduced distance equality constraints and have proposed a closed-form solution to the reconstruction problem. Due to the limitation of these constraints to recover sharp folds, we have shown how to replace them with distance inequalities, which yield a convex optimization problem.

In the future, we intend to study the use of our models, and potentially of our constraints to remove the requirement of a reference image. In [36], we started investigating this problem under the assumption that the surface remains locally planar. While this assumption is valid for smoothly deforming surfaces, such as the one of Fig. 18, it is not for sharp folds such as the one that appears in Fig. 19. Handling those, will require generalizing that approach.

Furthermore, we also intend to study the use of sources of information other than correspondences. In particular, the use of shading and silhouettes would give additional cues that could paliate the lack of texture. Ultimately, we hope such cues could be formulated in a similar convex optimization framework as our current approach.

References

[1] H. Aanaes and F. Kahl. Estimation of deformable structure and motion. In *Vision and Modelling of Dynamic Scenes Workshop*, 2002.

[2] A. Bartoli and S.I. Olsen. A Batch Algorithm For Implicit Non-Rigid Shape and Motion Recovery. In *ICCV Workshop on Dynamical Vision*, Beijing, China, October 2005.

[3] K. S. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popovic, and S. M. Seitz. Estimating cloth simulation parameters from video. In *ACM Symposium on Computer Animation*, 2003.

[4] V. Blanz and T. Vetter. A Morphable Model for The Synthesis of 3–D Faces. In *ACM SIGGRAPH*, pages 187–194, Los Angeles, CA, August 1999.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

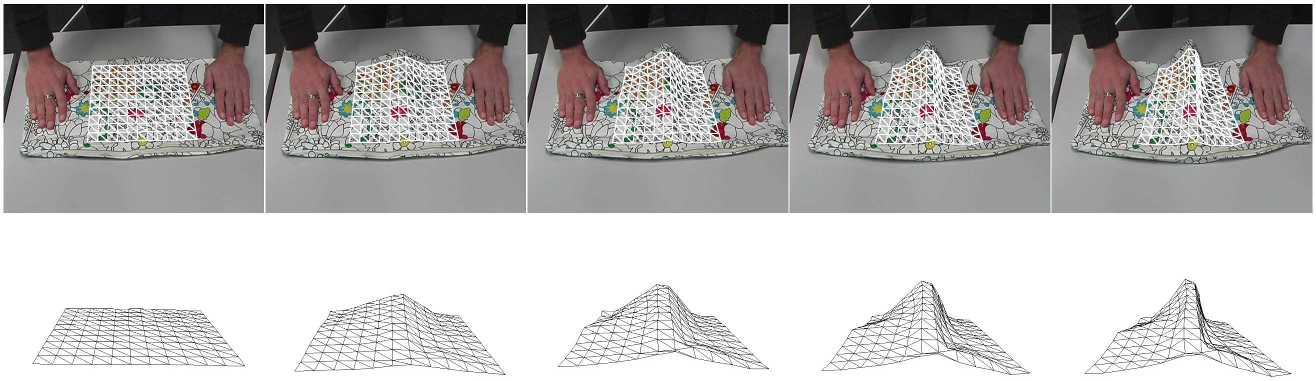


Fig. 19. Reconstructing a sharp fold in a piece of cloth. From top to bottom: Mesh recovered using equality constraints overlaid on the input image, side view of that mesh.

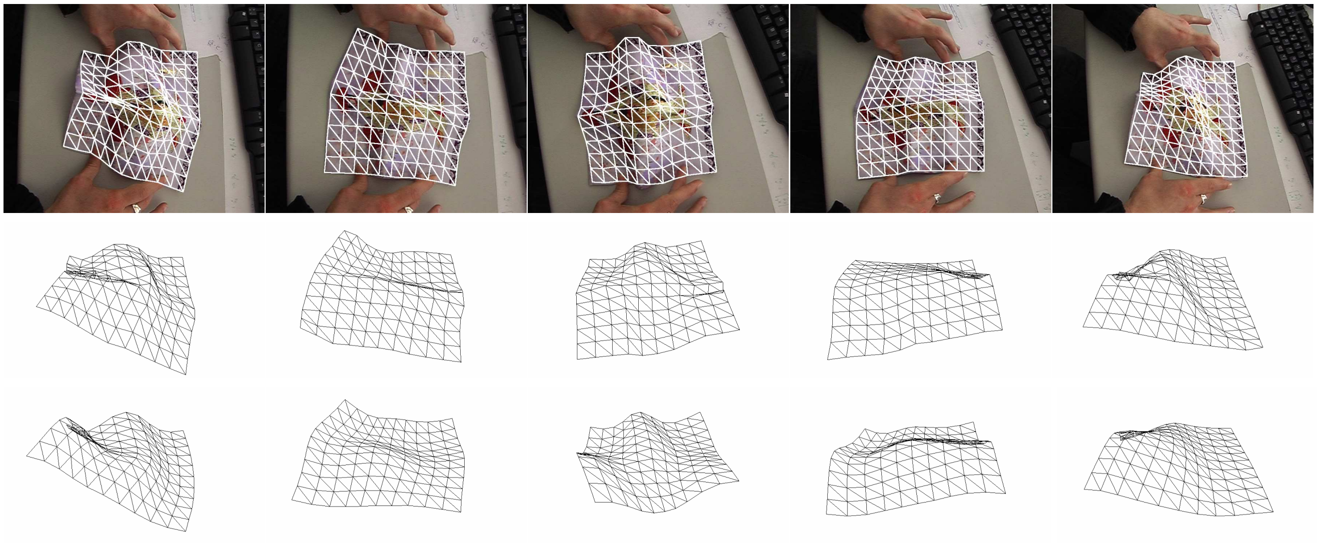


Fig. 20. Reconstruction of a deforming cloth. From top to bottom: Mesh recovered using inequality constraints overlaid on the image, side view of that mesh, side view of the mesh recovered using equality constraints. As in the synthetic case, using equality constraints results in oversmoothing whereas using inequalities does not.

[6] M. Brand. Morphable 3d models from video. *Conference on Computer Vision and Pattern Recognition*, 2001.

[7] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *Conference on Computer Vision and Pattern Recognition*, 2000.

[8] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[9] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In *European Conference on Computer Vision*, pages 484–498, Freiburg, Germany, June 1998.

[10] A. Ecker, A. D. Jepson, and K. N. Kutulakos. Semidefinite programming heuristics for surface reconstruction ambiguities. In *European Conference on Computer Vision*, Marseille, France, October 2008.

[11] N.A. Gumerov, A. Zandifar, R. Duraiswami, and L.S. Davis. Structure of Applicable Surfaces from Single Views. In *European Conference on Computer Vision*, Prague, May 2004.

[12] G. E. Hinton. Products of experts. In *International Conference on Artificial Neural Networks (ICANN)*, pages 1–6, 1999.

[13] N. D. Lawrence. Gaussian Process Models for Visualisation of High Dimensional Data. In *Neural Information Processing Systems*. MIT Press, Cambridge, MA, 2004.

[14] J. Liang, D. DeMenthon, and D. Doermann. Flattening curved documents in images. In *Conference on Computer Vision and Pattern Recognition*, pages 338–345, 2005.

[15] X. Llado, A. Del Bue, and L. Agapito. Non-rigid 3D Factorization for Projective Reconstruction. In *British Machine Vision Conference*, Oxford, UK, September 2005.

[16] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 20(2):91–110, 2004.

[17] I. Matthews and S. Baker. Active Appearance Models Revisited. *International Journal of Computer Vision*, 60:135–164, November 2004.

[18] T. McInerney and D. Terzopoulos. A Finite Element Model for 3D Shape Reconstruction and Nonrigid Motion Tracking. In *International Conference on Computer Vision*, pages 518–523, Berlin, Germany, 1993.

[19] D. Metaxas and D. Terzopoulos. Constrained deformable superquadrics and nonrigid motion tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.

[20] F. Moreno-Noguer, V. Lepetit, and P. Fua. Accurate Non-Iterative $O(n)$ Solution to the PnP Problem. In *International Conference on Computer Vision*, Rio, Brazil, October 2007.

[21] C. Nastar and N. Ayache. Frequency-based nonrigid motion analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11), November 1996.

[22] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:715–729, 1991.

[23] M. Perriollat and A. Bartoli. A quasi-minimal model for paper-like surfaces. In *BenCos Workshop at CVPR'07*, 2007.

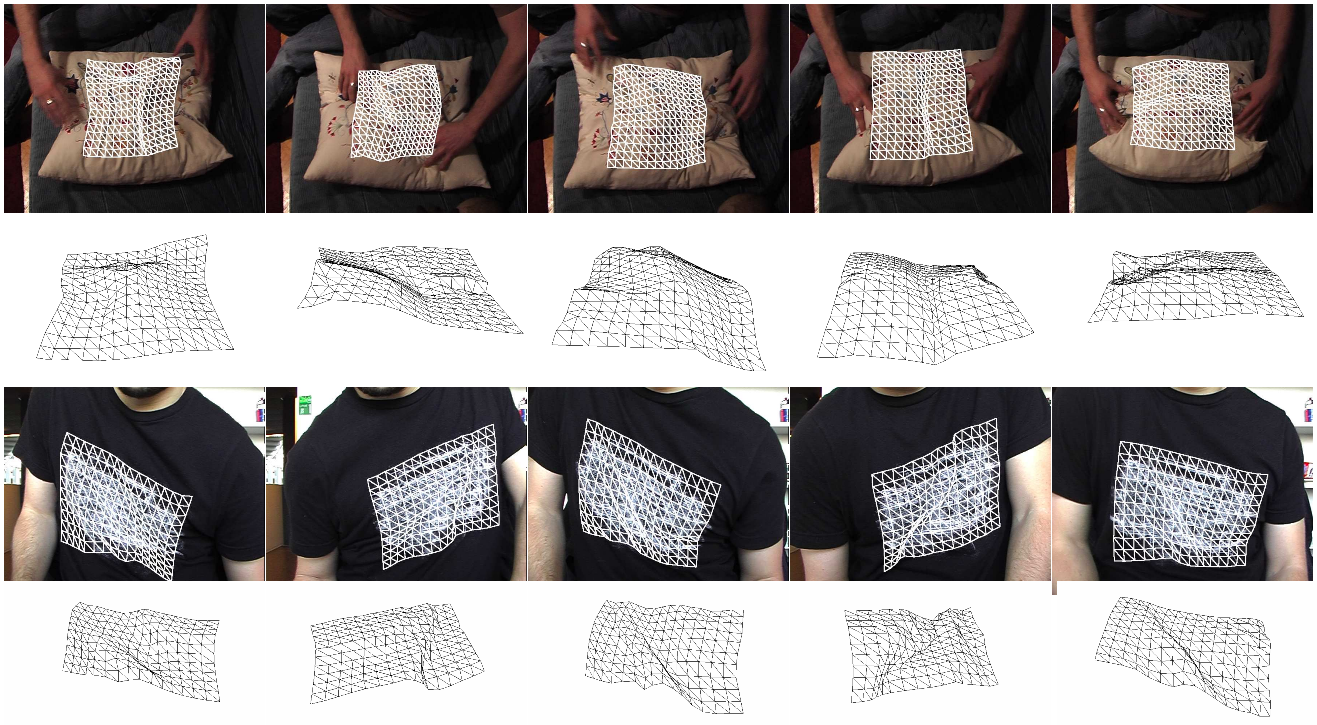
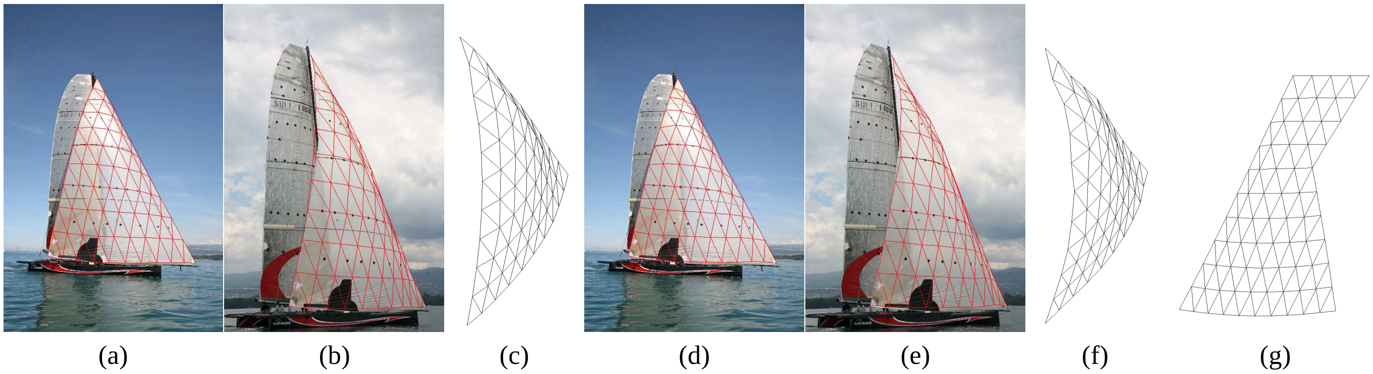


Fig. 21. We recovered several complex deformations of other cloth materials. First and third rows: Mesh recovered using inequality constraints overlaid on the original image. Second and fourth rows: Same mesh seen from a different viewpoint.



(a) (b) (c) (d) (e) (f) (g)

Fig. 22. Reconstruction of a triangular sail. (a,b) Shapes recovered with equality constraints overlaid on two original images. (c) Side view of the surface in (b). (d,e) Shapes recovered with distance inequalities overlaid on two original images. (f) Side view of the surface in (e). (g) Assembling local models to cover the entire surface required introducing additional vertices and facets. Note that they do not affect the reconstructions since they contain no correspondences.

[24] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. In *British Machine Vision Conference*, 2008.

[25] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. In *Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 2009.

[26] M. Salzmann, V. Lepetit, and P. Fua. Deformable Surface Tracking Ambiguities. In *Conference on Computer Vision and Pattern Recognition*, Minneapolis, MI, June 2007.

[27] 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*, Marseille, France, October 2008.

[28] M. Salzmann, J. Pilet, S. Ilić, and P. Fua. Surface Deformation Models for Non-Rigid 3–D Shape Recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1481–1487, February 2007.

[29] M. Salzmann, R. Urtasun, and P. Fua. Local deformation models for monocular 3d shape recovery. In *Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008.

[30] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, 1999.

[31] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-Rigid Structure from Locally Rigid Motion. In *Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 2010.

[32] D. Terzopoulos, J. Platt, A. Barr, and K. Fleicher. Elastically Deformable Models. *ACM SIGGRAPH*, 21(4):205–214, 1987.

[33] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3d shape from 2d motion. In *Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, 2003.

[34] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008.

[35] L. V. Tsap, D. B. Goldgof, and S. Sarkar. Nonrigid motion analysis based on dynamic refinement of finite element models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(5):526–543, 2000.

[36] R. Urtasun, D. Fleet, A. Hertzman, and P. Fua. Priors for people tracking

from small training sets. In *International Conference on Computer Vision*, Beijing, China, October 2005.

[37] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-Free Monocular Reconstruction of Deformable Surfaces. In *International Conference on Computer Vision*, Kyoto, Japan, October 2009.

[38] R. Vidal and R. Hartley. Perspective nonrigid shape and motion recovery. In *European Conference on Computer Vision*, Marseille, France, October 2008.

[39] J Xiao and T. Kanade. Uncalibrated perspective reconstruction of deformable structures. In *International Conference on Computer Vision*, 2005.

Appendix: Probabilistic Interpretation

In Section 5, we took the deformation energy of a mesh to be the sum of deformation energies over individual and overlapping patches. In probabilistic terms, this means that we compute the likelihood of a specific 3D shape as the product of the likelihood of its component patches. Since the patches share vertices, there are not independent from each other and it is therefore not completely obvious why this would result in the effective regularizer that our results show it to be. In this appendix, we provide empirical evidence as to why this is indeed the case.

To this end, we used motion capture data similar to what we used in Section 7.1. It was acquired by sticking 3mm wide hemispherical reflective markers on a rectangular surface and deforming it arbitrarily in front of six infrared Vicon^{TM} cameras that reconstruct the 3D positions of individual markers. We did this both for a $9x7$ grid of markers on a piece of cloth and a $9x9$ grid of markers on a piece of cardboard, the latter being of course much stiffer than the former. Let $\mathbf{X}^t = [x_1, y_1, z_1, ..., x_{P\times Q}, y_{P\times Q}, z_{P\times Q}]^T$ be the vector of the corresponding concatenated coordinates acquired at time t, with $P = 7$ and $Q = 9$ for the cloth and $P = 9$ and $Q = 9$ for the cardboard. In this manner, we acquired several thousand $\mathbf{X}^t$ vectors for each. The left column of Fig. 23 depicts the corresponding normalized covariance matrices and the right column their inverses, known as the *precision* matrices.

In this figure, dark red represents positive values, dark blue negative values, and light blue values close to zero. Therefore, if one treats these small values as truly being zero, the $\mathbf{P}$ precision matrices only have a few non zero diagonals for materials as different as cloth and cardboard. This is significant because, assuming that the $\mathbf{X}^t$ vectors are normally distributed, the likelihood of an arbitrary $\mathbf{X}$ vector can be estimated as

$$P(\mathbf{X}) \propto exp(-\mathbf{X}^T\mathbf{P}\mathbf{X}) \ . \tag{25}$$

Because closer examination of the $\mathbf{P}$ matrix reveals that its non-zero diagonals correspond to interactions between neighboring mesh vertices, this means that the likelihood of Eq. 25 can be rewritten as

$$P(\mathbf{X}) \propto \prod_i exp(-\mathbf{X}_i^T\mathbf{P_i}\mathbf{X}_i) \ , \tag{26}$$

where the $\mathbf{X}_i$ are the coordinates of the vertices of square patches such as those introduced in Section 5.1. $\log(P(\mathbf{X}))$ is therefore close to being a sum of terms computed over individual patches, which constitutes empirical evidence that our energy formulation is true to reality.

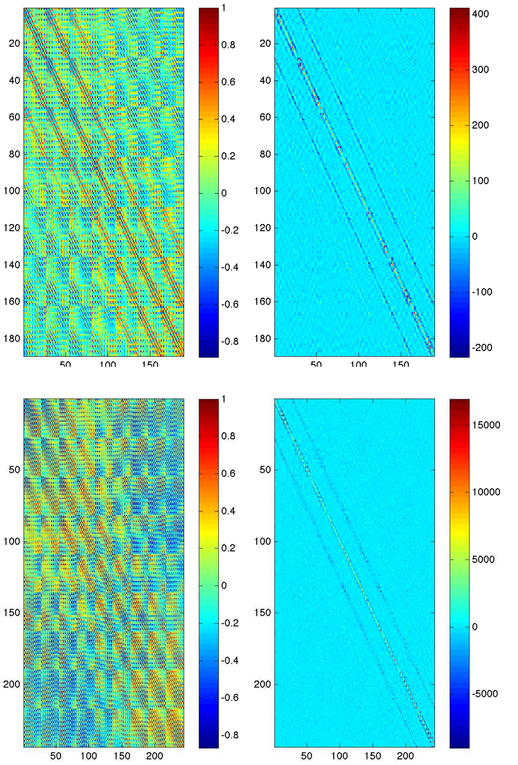


Fig. 23. Top row: Normalized covariance and precision matrices for the cloth data. Bottom row: The same matrices for the cardboard data. Note that the precision matrices are clearly banded if one treats the light blue areas as being zeros.

Mathieu Salzmann received his B.Sc and M.Sc degrees in computer science in 2004 from EPFL (Swiss Federal Institute of Technology). He obtained his PhD degree in computer vision in 2009 from EPFL. He then joined the International Computer Science Institute and the EECS departement at UC Berkeley as a postdoctoral fellow. Recently, he joined TTI Chicago as a Research Assistant Professor. His research interests include non-rigid shape recovery, human pose estimation, object recognition, and optimization techniques for computer vision.

Pascal Fua received the engineering degree from the Ecole Polytechnique, Paris, in 1984 and the PhD degree in computer science from the University of Orsay in 1989. He joined EPFL (Swiss Federal Institute of Technology) in 1996, where he is now a professor in the School of Computer and Communication Science. Before that, he worked at SRI International and at INRIA Sophia-Antipolis as a computer scientist. His research interests include shape modeling and motion recovery from images, human body modeling, and optimization-based techniques for image analysis and synthesis. He has (co)authored more than 150 publications in refereed journals and conferences. He has been an associate editor of the IEEE Transactions for Pattern Analysis and Machine Intelligence and has been a program committee member and an area chair of several major vision conferences.