Stochastic Exploration of Ambiguities for Non-Rigid Shape Recovery

Francesc Moreno-Noguer and Pascal Fua, *IEEE Fellow*

Abstract—Recovering the 3D shape of deformable surfaces from single images is known to be a highly ambiguous problem because many different shapes may have very similar projections. This is commonly addressed by restricting the set of possible shapes to linear combinations of deformation modes and by imposing additional geometric constraints. Unfortunately, because image measurements are noisy, such constraints do not always guarantee that the correct shape will be recovered. To overcome this limitation, we introduce a stochastic sampling approach to efficiently explore the set of solutions of an objective function based on point correspondences. This allows to propose a small set of ambiguous candidate 3D shapes and then use additional image information to choose the best one. As a proof of concept, we use either motion or shading cues to this end and show that we can handle a complex objective function without having to solve a difficult non-linear minimization problem. The advantages of our method are demonstrated on a variety of problems including both real and synthetic data.

Index Terms—Deformable surfaces, Monocular shape estimation.

1 Introduction

For the purpose of single view deformable 3D shape reconstruction, approaches that rely on purely geometric constraints can return incorrect answers because there are often many different shapes that obey, or nearly obey these constraints, while producing very similar projections.

For example, it has been shown that non-rigid 3D shape could be recovered from even single images provided that enough correspondences can be established between that image and one in which the surface's shape is already known [25], [27], [37]. Yet, while effective, these techniques return one single reconstruction without accounting for the fact that several plausible shapes could produce virtually the same projection and therefore be indistinguishable on the basis of correspondences and geometry alone. In practice, as shown in Fig. 1, disambiguation is only possible using additional image information.

In this paper, we propose a generic approach to exploring the space of feasible solutions, which is grounded on the theory of uncertainty propagation and stochastic search. More specifically, we represent shape deformations as a weighted sum of deformation modes and relate uncertainties on the location of point correspondences to uncertainties on the modal weights. This lets us explore the space of modes using a stochastic sampling strategy and select a small number of ambiguous solutions, which correspond to 3D non-rigid shapes such as those shown in Fig. 1. As a proof of concept, we then propose two different approaches to disambiguation:

- **Exploiting shading information.** The best 3D shape is chosen among the candidates generated in this manner using shading information, both when the light sources are distant and when they are nearby. The latter is particularly significant because exploiting nearby light sources would involve solving a difficult non-linear minimization problem if we did not have a reliable way to generate 3D shape hypotheses. In our examples, this is all the more true since the lighting parameters are initially unknown and must be estimated from the images.
- **Enforcing temporal consistency.** Assuming that a video sequence is available, we will exploit three-frame sequences to pick the set of candidate 3D shapes that provides the most temporally consistent motion. Note that in contrast to traditional tracking and non-rigid shape from motion approaches [5], [24], [31], [35], we do not enforce temporal consistency across the whole sequence and, therefore, do not require points to be tracked across many images.

We show that both these approaches outperform state-of-the-art non-rigid shape recovery methods [23], [28], and allow disambiguating 3D shapes without having to learn complex mappings or tracking a large number of frames.

In short, the contribution of this paper, which extends an earlier conference paper [22], is an approach to avoiding being trapped in the local minima of a potentially complicated objective function by efficiently exploring the solution space of a simpler one. As a result, we only

- *F. Moreno-Noguer is with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, 08028, Spain.*
  E-mail: see http://www.iri.upc.edu/people/fmoreno/
- *P. Fua is with the Computer Vision Laboratory, EPFL, Lausanne, 1015, Switzerland.*
  E-mail: see http://cvlab.epfl.ch/∼fua/

This work has been partially funded by the Spanish Ministry of Science and Innovation under CICYT project PAU+ DPI2011-27510; by MIPRCV Consolider Ingenio 2010 CSD2007-00018; by the EU project GARNICS FP7-247947 and by the Swiss National Science Foundation.

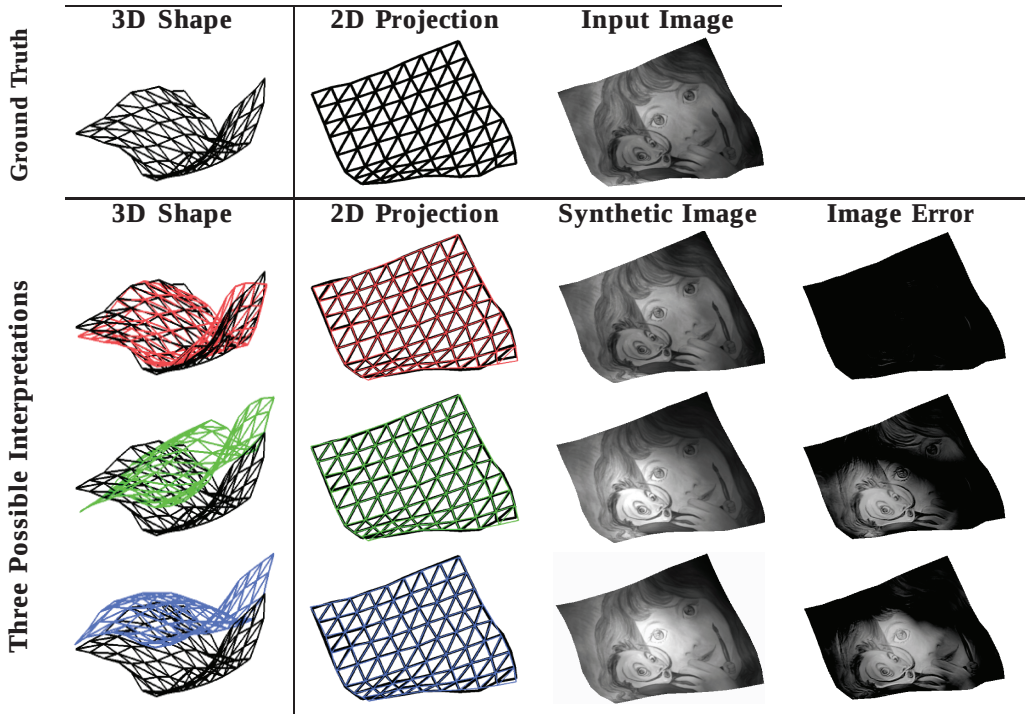

Fig. 1. Handling 3D shape ambiguities. **First Row:** Image of a surface lit by a nearby light source and the corresponding ground truth surface. **Three other Rows:** In each one, a different candidate surface proposed by our algorithm is shown in color. The corresponding projection and synthesized image given automatically estimated lighting parameters are shown in the middle columns. As can be seen, its projection is very similar, even though its shape may be very different from the original one. In other words, the candidates cannot be distinguished based on reprojection error alone. However, when comparing the true and synthesized images, it becomes clear that the correct shape is the one at the second row.

need to evaluate the full objective function for a few selected configurations, which implies we could use a very discriminating and expensive one if necessary.

2 Related Work

Single-view 3D reconstruction of non-rigid surfaces has been extensively studied over the years. It is known to be a highly under-constrained problem that cannot be solved without *a priori* knowledge.

A typical approach to introducing such knowledge is to use deformation models, either physically inspired ones [7], [20], [21], [32] or learned from training data [3], [4], [5], [8], [29]. Surface deformations are then expressed as weighted sums of modes and retrieving shape entails estimating the modal weights by minimizing an image-based objective function. However, since such functions usually have many local minima, a good initialization is required.

There have also been recent attempts at recovering the shape of inextensible surfaces without explicitly using a deformation model. Some approaches are specifically designed for applicable surfaces such as sheets of paper [12], [17]. Others make use of local inextensibility constraints [11], [25], and are applicable to many materials that do not perceptibly shrink or stretch as they deform. However, while these are attractive approaches, using only inextensibility constraints is only effective for relatively small deformations.

Other methods use local rigidity constraints in conjunction with deformation models and achieve shape-recovery either in closed form [28] or by solving a convex optimization problem [27], and thus, eliminate the need for an initialization. To this end, they require 2D point correspondences between the image in which one wishes to compute the shape and one in which it is already known. However, as will be shown in the following sections, small inaccuracies in the correspondences can still result in erroneous reconstructions.

The method proposed in this paper builds on the formalism introduced in [28] to return not a single solution but a representative set of *all* possible solutions and then uses additional information to decide which one is best. In this paper, we use shading and motion but any image cue could have been used instead.

Of course, some of the core elements of our approach have appeared before in the literature. For instance, a large number of recent methods, such as [1], [10], [33], [34], have been proposed to merge geometric and shading cues into a common framework. However, these techniques, unlike ours, involve multiple iterative pro-

	Shape # 1	Shape # 2	Shape # 3
Reconst. Error (mm)	0.82	4.25	5.35
Reproj. Error (pix)	1.92	1.87	1.93
Inextens. Error (mm)	4.00	4.27	3.97

TABLE 1
Mean reconstruction, reprojection and inextensibility errors for the candidate shapes of Fig. 1. Note that, although Shape #1 violates edge-length constraints slightly more than Shape #3, it still is the reconstruction closest to the ground truth by far.

cesses that require reasonable good initial estimates. An exception is the algorithm of [23] that solves for shape in closed form but is only applicable to Lambertian surfaces lit by a distant point light source. In addition, shape ambiguities have also been discussed before, but mostly in the context of rigid surface estimation [6], [26], [30]. In the non-rigid case, [11] is the only paper we know of that discusses shape ambiguities. This discussion, however, is limited to ambiguities produced by the concave / convex reversal. By contrast, we propose here a stochastic sampling exploration method that is effective to handle continuous ambiguities.

3 Exploring the Solution Space

Let us now assume that we are given a *reference image* in which the 3D shape is known and a set of 2D point correspondences between this reference image and an *input image* in which the shape is unknown. In order to compute the unknown 3D shape we will use the same formalism as in [28] and will represent the deformations in terms of a weighted sum of modes. We will then seek for the weights that minimize the reprojection error while preserving the distances in local neighborhoods. However, and in contrast to [28], we will not just retain a single solution as the correct shape is not always the one that minimizes the above geometric criterion. This is shown in Table 1 for the 3D surfaces of Fig. 1, and responds to the fact that since in practice the 3D-to-2D correspondences are not infinitely accurate, small amounts of reprojection error may result in large changes in 3D shape.

To avoid this problem, instead of picking the best set of weights according uniquely to the geometric criterion, we will fit a Gaussian distribution to those that correspond to acceptable projections. This will let us to exhaustively sample the sets of weights that also preserve local distances, and will typically result in approximately one hundred candidate shapes per image, among which the best will be picked using additional sources of image information. In Section 4 we will then use either shading cues or temporal consistency constraints for this purpose.

3.1 Problem Formulation

In this section we will show that given a set of 3D-to-2D correspondences, the problem of recovering non-rigid shape can be found as the solution of a linear system.

As depicted in Fig. 2, we represent non-rigid surfaces as triangulated meshes with n_v vertices $\mathbf{v}_i$ concatenated in a vector $\mathbf{x} = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_{n_v}^\top]^\top$. We model shape deformations as weighted sums of n_m deformation modes $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_{n_m}]$. We write

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{n_m} \alpha_i \mathbf{q}_i = \mathbf{x}_0 + \mathbf{Q}\boldsymbol{\alpha} \,, \tag{1}$$

where $\mathbf{x}_0$ is a mean shape and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{n_m}]^\top$ are unknown weights that define the current shape. The deformation modes were obtained by applying Principal Component Analysis over a set of deformed shapes, randomly generated for the synthetic experiments, or learned from training data in the real experiments.

As in [23], [27], we treat a correspondence between a 2D point $\mathbf{u}_i^{\text{ref}}$ in the reference image and a 2D point $\mathbf{u}_i$ in the input image as a 2D-to-3D correspondence between $\mathbf{u}_i$ and $\mathbf{p}_i^{\text{ref}}$, the 3D point on the shape in its reference configuration that projects at $\mathbf{u}_i^{\text{ref}}$.

We then express the coordinates of $\mathbf{p}_i$, the unknown position of the point $\mathbf{p}_i^{\text{ref}}$ in the deformed mesh, in terms of the barycentric coordinates of the face to which belongs

$$\mathbf{p}_i = \sum_{j=1}^{3} a_{ij} \mathbf{v}_j^{[i]} \,, \tag{2}$$

where the a_{ij} are the barycentric coordinates and the $\mathbf{v}_j^{[i]}$ are the unknown nodes we seek to retrieve. Note that the barycentric coordinates let us write the location of the point coordinates in terms of the mesh vertices. We compute them from the position $\mathbf{p}_i^{\text{ref}}$ of the points in the reference mesh, and remain the same for the deformed mesh, as we assume the mesh is inextensible.

Additionally, we assume the matrix $\mathbf{A}$ of internal camera parameters to be known and that the 3D points are expressed in the camera reference frame. Therefore, the fact that $\mathbf{p}_i$ projects at $\mathbf{u}_i$ implies that

$$w_i \begin{bmatrix} \mathbf{u}_i \\ 1 \end{bmatrix} = \mathbf{A}\mathbf{p}_i = \begin{bmatrix} \mathbf{A}_{2\times 3} \\ \mathbf{a}_3^\top \end{bmatrix} \mathbf{p}_i \,, \tag{3}$$

where w_i is a projective scalar, $\mathbf{A}_{2\times 3}$ are the first two rows of $\mathbf{A}$ and $\mathbf{a}_3^\top$ the last one. Since from the last row we have that $w_i = \mathbf{a}_3^\top \mathbf{p}_i$, we can write

$$\left(\mathbf{u}_i \mathbf{a}_3^\top - \mathbf{A}_{2\times 3}\right) \mathbf{p}_i = \mathbf{0} \,. \tag{4}$$

By using Eq. 2 we can represent $\mathbf{p}_i$ as a function of the mesh vertices. Thus, for each 3D-to-2D correspondence, Eq. 4 provides 2 linear constraints on $\mathbf{x}$. n_c such correspondences yield $2n_c$ constraints which can be written as a linear system

$$\mathbf{M}\mathbf{x} = \mathbf{0} \,, \tag{5}$$

where $\mathbf{M}$ is a $2n_c \times 3n_v$ matrix obtained from the known values $\mathbf{u}_i$, $\mathbf{A}$, and a_{ij}. Injecting the modal description of Eq. 1 then yields

$$\mathbf{M}\mathbf{Q}\boldsymbol{\alpha} + \mathbf{M}\mathbf{x}_0 = \mathbf{0} \,, \tag{6}$$

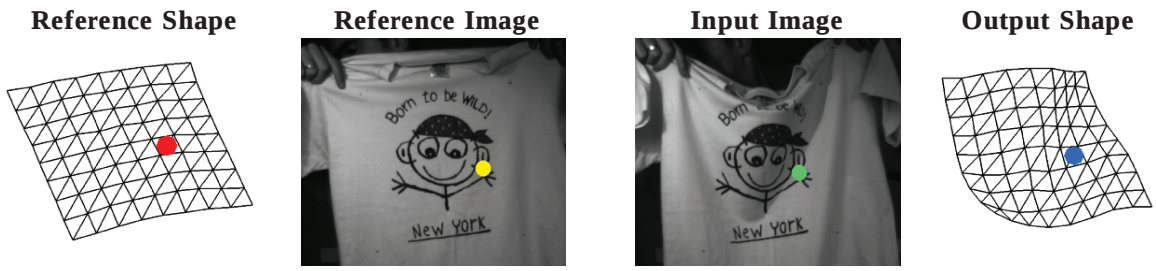

Fig. 2. Estimating non-rigid shape from 3D-to-2D correspondences. We assume we are given a set of 2D-to-2D point correspondences between a reference image and the input image. In addition, we assume the reference image to be registered to a known shape, and thus, the correspondences are in fact from 3D-to-2D. Our goal, is to retrieve the output shape from these matches. The colored dots represent corresponding points for the reference, input and output configurations. We denote the 3D position of the point in the reference shape (red dot) by $\mathbf{p}^{\text{ref}}$. The 2D projection (yellow dot) of this point on the reference image is denoted by $\mathbf{u}^{\text{ref}}$. Its 2D correspondence (green dot) on the input image is written by $\mathbf{u}$, and $\mathbf{p}$ is its unknown 3D position (blue dot).

such that any set of weights $\boldsymbol{\alpha}$ that is a solution of it, corresponds to a 3D surface that projects at the right place.

3.2 Proposing Candidate Shapes

Since correspondences $\{\mathbf{p}_i^{\text{ref}}, \mathbf{u}_i\}$ are potentially noisy, the simplest way to solve Eq. 6 is in the least-squares sense. This, however, may not result in a satisfactory answer because $\mathbf{MQ}$ is an ill-conditioned matrix with several small eigenvalues [23], [28]. As a result, even when there are many correspondences, small changes in the exact correspondence locations, and therefore in the coefficients of $\mathbf{M}$, can result in very large changes of the resulting $\boldsymbol{\alpha}$ values. In other words, many different sets of $\boldsymbol{\alpha}$ weights can result in virtually the same projection. In [28], this is addressed by choosing the weights that best preserve the lengths of the mesh edges. However, as shown by Table 1, this does not necessarily yield the best answer.

Therefore, in this paper, instead of choosing the best set of weights based on geometric considerations alone we have devised a way to quickly propose a restricted set of candidate solutions among which the best can be chosen using additional sources of image information, as will be done in Section 4. To this end, we first derive an analytical expression of the solution space as a function of the 2D input data statistics. We then efficiently sample this space and keep the best samples in terms of both minimizing reprojection errors and preserving local distances.

3.2.1 Gaussian Representation of the Solution Space

The $\boldsymbol{\alpha}$ weights we seek can be computed as the least-squares solution of Eq. 6:

$$\boldsymbol{\alpha} = (\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{b} \ , \tag{7}$$

where $\mathbf{B} = \mathbf{MQ}$ is a $2n_c \times n_m$ matrix, and $\mathbf{b} = -\mathbf{M}\mathbf{x}_0$ is a $2n_c$ vector. The components of $\mathbf{B}$ and $\mathbf{b}$ are linear functions of the known parameters $\mathbf{u}_i$, $\mathbf{Q}$, $\mathbf{A}$ and a_{ij}. We have seen that this solution may not, in fact, be the right one because $\mathbf{B}$ is ill-conditioned and solving the system in the least-squares sense magnifies small inaccuracies in the correspondences. We can nevertheless exploit the expression of Eq. 7 to model where to look for other potential solutions as follows.

Let us assume that the estimated correspondence locations are normally distributed around their true locations. Injecting the corresponding $2n_c \times 2n_c$ diagonal covariance matrix $\boldsymbol{\Sigma}_{\mathbf{u}}$ into Eq. 7 means that the $n_m \times n_m$ covariance matrix for the $\boldsymbol{\alpha}$ weights can be written as

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \mathbf{J}_\beta \boldsymbol{\Sigma}_{\mathbf{u}} \mathbf{J}_\beta^\top \ , \tag{8}$$

where $\mathbf{J}_\beta$ is the $n_m \times 2n_c$ Jacobian of $(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{b}$ with respect to the 2D correspondence coordinates. Thus

$$\mathbf{J}_\beta = \frac{\partial(\mathbf{B}^\top\mathbf{B})^{-1}}{\partial \mathbf{u}}\mathbf{B}^\top\mathbf{b} + (\mathbf{B}^\top\mathbf{B})^{-1}\frac{\partial \mathbf{B}^\top\mathbf{b}}{\partial \mathbf{u}} \ , \tag{9}$$

which can be computed analytically considering that

$$\frac{\partial(\mathbf{B}^\top\mathbf{B})^{-1}}{\partial \mathbf{u}} = -(\mathbf{B}^\top\mathbf{B})^{-1}\frac{\partial(\mathbf{B}^\top\mathbf{B})}{\partial \mathbf{u}}(\mathbf{B}^\top\mathbf{B}) \ . \tag{10}$$

We can therefore represent the family of 3D surfaces whose projections are close to the one that minimizes the reprojection error as being normally distributed around $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, the least squares solution of Eq. 6, with covariance $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ of Eq. 8. Note that, because $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ is the solution of an ill-conditioned system, it is an unreliable estimate of the distribution's center. We could have improved the system's conditioning by adding a damping term, but this would have amounted to arbitrarily constraining the norm of $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$. Instead, as discussed in the next section, we use a stochastic sampling mechanism to explore different possible values of $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$.

3.2.2 Efficiently Exploring the Solution Space

To create a set of plausible 3D shapes whose projection are acceptably close to the correct one, we first define a search region in the n_m-dimensional space of the modal weights. We then explore it using a sampling approach which is based on an Evolution Strategy.

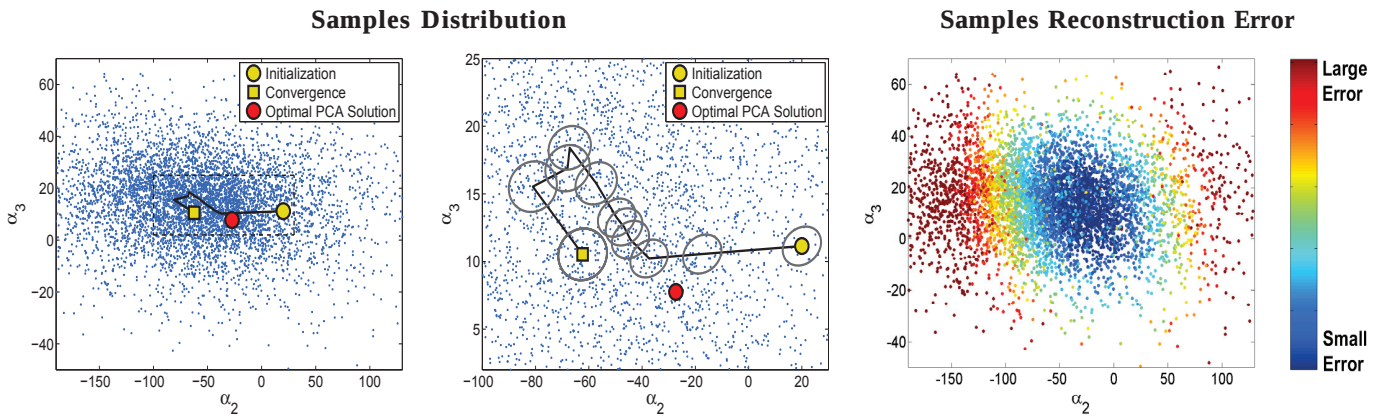

Fig. 3. Efficient exploration of the solution space. The left-most figure shows the distribution of samples on the modal weights space. For visualization purposes, only two of the n_m dimensions are represented. In addition to the individual samples, the graph depicts the path followed by the mean $\mu_{\boldsymbol{\alpha}}$ for successive iterations, the initial and final configurations, and an optimal solution computed by directly projecting the ground-truth shape onto the deformation modes. The evolution path is estimated using the CMA algorithm by minimizing an objective function of the reprojection and inextensibility errors. The area inside the dashed rectangle is magnified in the middle image, and shows a detail of the updated mean and covariance matrix. Note that although the mean evolution path does not end up close to the optimal solution, some of the samples accumulated through the process lie very close, and are potentially good solutions. The right-most figure depicts the reconstruction error for each of the samples, color-coded according to the right bar.

Given the normal distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ introduced above, we take the search region as the set of $\boldsymbol{\alpha}_i$ such that

$$(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) \leq \mathcal{M}^2 \ , \tag{11}$$

where $\mathcal{M}$ is a threshold chosen to achieve a specified degree of confidence. To compute its value we use the cumulative chi-squared distribution, which depends on the dimensionality of the problem. In our experiments, $n_m = 30$ modes were sufficient to capture all surface deformations. For this number of modes, setting $\mathcal{M} = 7$ yields a 98% level of confidence.

We could then explore this region by drawing random samples from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \mathcal{M}^2\boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$. However, as the $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ we use is unreliable, and both $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ are built on the basis of uniquely minimizing the reprojection error, we do not draw all samples at once. Instead, we propose an evolution strategy in which we draw successive batches by sampling from a multivariate gaussian whose mean and covariance are iteratively updated in order to fit an energy landscape that simultaneously minimizes reprojection and inextensibility errors. The adaptation procedure is inspired by the Covariance Matrix Adaption algorithm [14], an iterative random sampling method, which has been shown to be effective for optimizing non-linear objective functions. It includes the following steps:

1) Let $k \in \mathbb{N}$ denote the current iteration, and $\boldsymbol{\Lambda}$ the set of sample shapes which are accumulated throughout the process. Initially, at $k = 0$, we set the mean $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^k$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^k$ to the values estimated using Eqs. 7 and 8, respectively. $\boldsymbol{\Lambda}$ is initialized to an empty set.
2) We then draw n_s random samples $\{\tilde{\boldsymbol{\alpha}}_i^k\}_{i=1}^{n_s}$ from the distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^k, \mathcal{M}^2\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^k)$.
3) Each sample $\tilde{\boldsymbol{\alpha}}_i^k$ is assigned a weight π_i^k which simultaneously considers the reprojection and inextensibility errors:

$$\frac{1}{\pi_i^k} = \lambda_r \texttt{Repr_Err}(\tilde{\boldsymbol{\alpha}}_i^k) + \lambda_i \texttt{Inext_Err}(\tilde{\boldsymbol{\alpha}}_i^k) \ . \tag{12}$$

Since these errors are expressed in different units of measurement, we use λ_r and λ_i to give them similar orders of magnitude.
The reprojection and inextensibility errors above are computed as follows:
Let $\tilde{\mathbf{x}} = [\tilde{\mathbf{v}}_1^\top, \ldots, \tilde{\mathbf{v}}_{n_v}^\top]^\top$ be the shape corresponding to a sample $\tilde{\boldsymbol{\alpha}}$ in the modal weights space, and let $\{\tilde{\mathbf{u}}_i\}_{i=1}^{n_c}$ be the 2D projections of the 3D points for which correspondences $\mathbf{u}_i$ are available. We then define:

$$\texttt{Repr_Err}(\tilde{\boldsymbol{\alpha}}) = \sum_i^{n_c} \|\tilde{\mathbf{u}}_i - \mathbf{u}_i\| \tag{13}$$

$$\texttt{Inext_Err}(\tilde{\boldsymbol{\alpha}}) = \sum_{\{i,j\}\in\mathcal{I}} \|\tilde{l}_{ij} - l_{ij}^{\text{ref}}\| \ , \tag{14}$$

where $\tilde{l}_{ij}$ is the distance between two neighboring nodes $\tilde{\mathbf{v}}_i$ and $\tilde{\mathbf{v}}_j$, l_{ij}^{ref} is the distance between the same nodes in the reference configuration, and $\mathcal{I}$ represents the indices of neighboring nodes.
4) As in the CMA algorithm, the mean and covariance matrix are updated as follows:

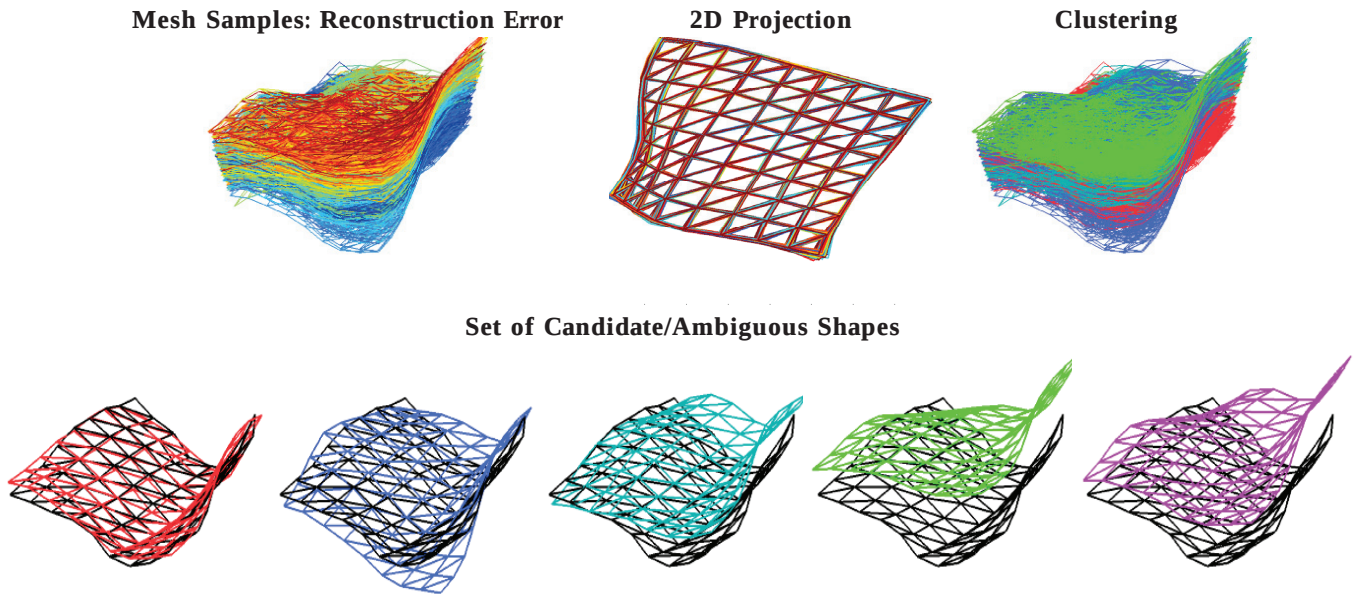

Fig. 4. Clustering the Shape Samples. **Top Row:** The left-most figure shows the meshes corresponding to the modal weight samples of Fig. 3. Again, the color of the meshes encodes the reprojection error. The figure in the middle shows the reprojection of the meshes on the image plane. Observe that their projection is very similar. The figure on the right shows the result of classifying the shape samples into a few clusters, where each cluster is represented by a different color. **Bottom Row:** The cluster centers are taken as the set of potential candidate shapes that best span the solution space. In practice the number of ambiguous shapes we estimate this way is around one hundred.

The mean vector $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{k+1}$ is estimated as a weighted average of the samples:

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{k+1} = \frac{\sum_{i=1}^{n_s} \pi_i^k \tilde{\boldsymbol{\alpha}}_i^k}{\sum_{i=1}^{n_s} \pi_i^k} \,. \tag{15}$$

The update of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{k+1}$ consists of three terms: a scaled covariance matrix of the previous step, a covariance matrix $\boldsymbol{\Sigma}_{\text{curr}}$ that estimates the variances of the best sampling points in the current generation, and a covariance matrix $\boldsymbol{\Sigma}_{\text{evol}}$ that exploits information of the correlation between the current and previous generations,

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{k+1} = \sigma^k \left[(1-\lambda_c-\lambda_e)\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^k + \lambda_c \boldsymbol{\Sigma}_{\text{curr}} + \lambda_e \boldsymbol{\Sigma}_{\text{evol}}\right] \,. \tag{16}$$

The parameters λ_c and λ_e are precomputed learning rates, and σ^k controls both the global scale of the distribution and the step size of the evolution path. For a discussion on how these parameters are chosen and further details of the CMA algorithm we refer the reader to [14], [15].

5) The best $n_{sb} << n_s$ samples with larger weights are retained and added to the set $\boldsymbol{\Lambda}$.
6) Steps 2) - 5) are repeated until a maximum number of iterations `Max_Iter` is reached or until the mean fitness error of Eq. 12 for all samples drops below a threshold.

In our experiments we used $n_s = 200$ random samples, $n_{sb} = 50$, and we set `Max_Iter` $= 200$, which lead to a maximum number of $200 \cdot 200 = 4 \times 10^4$ samples to explore the solution space, among which $200 \cdot 50 = 10^4$ were retained for further analysis. The parameters λ_r and λ_i were set to 0.8 and 0.2, respectively, and the parameters of the CMA algorithm were set to the default values suggested in [15].

The left image of Fig. 3 shows an example of the type of distribution we obtain with the proposed approach. Note that although the CMA converges relatively far from the optimal solution with minimal reconstruction error, some of the samples accumulated through the exploration process are in fact very good approximations. This illustrates the advantages of an approach like ours that simultaneously considers several plausible solutions, and not only one. Furthermore, the cost of obtaining this set of solution is very low, as it only requires evaluating Eq. 12, which may be done very efficiently.

In the following sections we will progressively apply more stringent, and more computationally demanding constraints to an ever decreasing number of samples, until obtaining one single solution.

3.2.3 *Clustering the Shape Samples*

By construction, all the samples generated above represent shapes that yield similar projections and only small violations of the length constraints. Furthermore, many of them yield almost undistinguishable 3D shapes. To further reduce their number, we therefore run a Gaussian-means clustering algorithm over all the accumulated samples in the space of the 3D coordinates. For this purpose we used [13], which is a variant of the k-means algorithm that automatically identifies the optimal number of clusters based on statistical tests

designed to check whether all the clusters follow a Gaussian distribution. These tests are controlled by means of a significance level parameter which we set to a very low value to favor over-segmentation, that is, to produce more clusters than absolutely necessary to avoid grouping shapes whose difference is statistically significant.

Finally, we take our set of candidates shapes to be the cluster centers. This whole process typically reduces the 10^4 samples of the previous stage to about one hundred. Fig. 4 shows a few candidate samples we obtain. Observe that even though they have very similar projections, their 3D shape is very different.

4 Using Additional Cues to Select the Best Candidate

Given a set of correspondences between the reference and the input images, the algorithm discussed in the previous section returns about 100 candidate 3D shapes that all project correctly in the input image and whose local distances have retained their original length. We will next show how to use additional image information to disambiguate and pick the best one.

In this paper we will either use shading or motion for this purpose. Note however, that these approaches are just meant to be a proof of concept, and many other image cues, such as silhouettes, texture or shadows, could be used.

4.1 Using Shading to Disambiguate

When using shading to disambiguate among several surface candidates, we consider two different cases. First, we assume the surface is lit by a *distant light source*, which is the situation envisioned in earlier works on monocular deformable surface reconstruction that use shading cues [23], [33], [34]. Second, we address the situation in which the surface is lit by a *nearby light source*. This is more difficult because the inverse of the changing distance to the light source has to be taken into account, which rules out approaches based on simple linear or quadratic constraints. In both cases, we do not assume the lighting parameters to be known *a priori* and estimate them from the candidate 3D shapes. As shown in Fig. 1, this lets us render the image we would see for any candidate shape, compare it to the real one, and select the best. To perform the rendering, we use ray-tracing and take into account visibility effects and shadows cast by the object on itself. Such non-local and non-linear phenomena are rarely taken into account by continuous optimization-based schemes because they result in highly complex energy landscapes and poor convergence. We now turn to the estimation of the lighting parameters in these two cases.

4.1.1 Light Source at Infinity

Recall from Section 3.1, that we start from a set of correspondences between 3D surface points $\mathbf{p}_i^{\text{ref}}$ and 2D

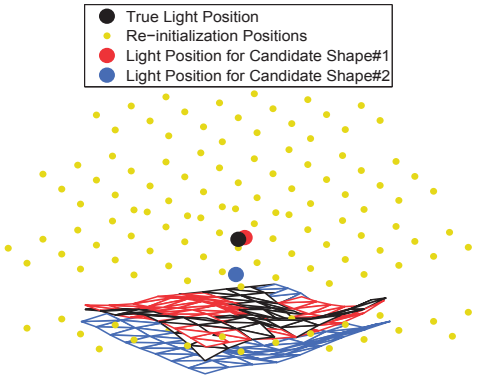

Fig. 5. Using shading to disambiguate under the assumption of a nearby light source. **Top:** The red (Shape #1) and blue (Shape #2) meshes are two possible interpretations of the ground truth mesh in black. To pick the best one, we estimate in each case the light source position through an optimization procedure which is re-initialized on all the yellow dots. It can be clearly seen that the light source position estimated by Shape #1 is more accurate. **Bottom:** Since the true light source position is unknown, we pick the best shape by using the estimated light sources to synthesize the input image and compute the error. The intensity error for Shape #1 is considerably smaller, and hence this is the chosen shape. Note that our approach allows to simultaneously estimate shape and lighting parameters.

image points $\mathbf{u}_i$ in the input image with intensity I_i. For each point i, we also know that $\mathbf{p}_i^{\text{ref}}$ projects at $\mathbf{u}_i^{\text{ref}}$ in the reference image and has intensity I_i^{ref}. In practice, we acquire the reference image under diffuse lighting so that, assuming the surface to be Lambertian, we can take the albedo ρ_i of $\mathbf{p}_i^{\text{ref}}$ to be I_i^{ref}. In the remainder of this Section, let $\mathbf{p}_i$ denote the 3D coordinates of the 3D surface points in the candidate shapes. For each candidate shape, these $\mathbf{p}_i$ are recomputed using the barycentric coordinates, which are the same for all candidates, to average the 3D vertex coordinates of the facets they belong to.

Assuming a distant light source parameterized by its unit direction $\mathbf{l}$ and power L, we can write $I_i = \rho_i L(\mathbf{l} \cdot \mathbf{n}_i)$, where $\mathbf{n}_i$ is the surface normal at $\mathbf{p}_i$, which may be estimated from the $\mathbf{v}_i$ vertex coordinates. Grouping these equations for all n_c correspondences yields

$$\mathbf{I}_\rho = \mathbf{N}\mathbf{L} \ , \tag{17}$$

where $\mathbf{I}_\rho = [I_1/\rho_1, \ldots, I_{n_c}/\rho_{n_c}]^\top$, $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_{n_c}]^\top$,

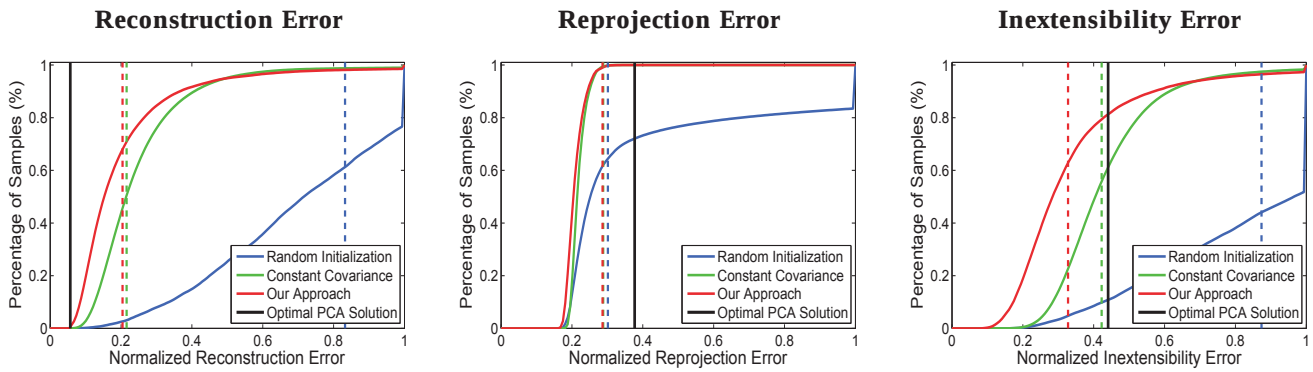

Fig. 6. Distribution of the reconstruction (left), reprojection (center) and inextensibility (right) errors for all the samples accumulated using three exploration strategies. The solid lines represent the percentage of samples –vertical axis– with a maximum level of error indicated by the horizontal axis. The dashed vertical lines correspond to the error of the mean vector $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{k}$ in Eq. 15 at the convergence of the exploration procedures. We also plot the optimal PCA solution, which is computed by projecting the ground truth shapes onto the deformation modes, and represents the best –minimal reconstruction error– approximation of the ground truth shape we could obtain.

and $\mathbf{L} = L \cdot \mathbf{l}$. Solving this system in the least-squares sense yields an estimation of $\mathbf{L}$, from which the light intensity and direction can be taken to be $L = \|\mathbf{L}\|$ and $\mathbf{l} = \mathbf{L}/L$.

4.1.2 Nearby Light Source

When considering a light source that is not located at infinity, the fact that the brightness decreases with the square of the distance must be taken into account. The image irradiance at $\mathbf{p}_i$ therefore becomes

$$I_i = \rho_i L \frac{\mathbf{l}_i \cdot \mathbf{n}_i}{\|\mathbf{p}_i - \mathbf{s}\|^2} \ , \tag{18}$$

where $\mathbf{l}_i = \frac{1}{\|\mathbf{p}_i - \mathbf{s}\|}(\mathbf{p}_i - \mathbf{s})$ and $\mathbf{s}$ is the position of the light source. $\mathbf{s}$ and L are estimated by minimizing

$$\sum_{i=1}^{n_c} \left| I_i - \rho_i L \frac{\mathbf{l}_i \cdot \mathbf{n}_i}{\|\mathbf{p}_i - \mathbf{s}\|^2} \right| \tag{19}$$

with respect to L and $\mathbf{s}$ using the nonlinear least-squares Matlab routine `lsqnonlin`. To avoid local minima, we define a sparse set of light positions $\{\tilde{\mathbf{s}}_j\}_{j=1}^{n_l}$ and use each one in turn to initialize the optimization. In our experiments, we used $n_l = 125$ light positions uniformly distributed within a hemisphere on top of the reference mesh. Its radius was taken to be sufficiently large to include all distances for which the nearby light assumption holds. Fig. 5 shows a simple example of this methodology.

Note that what makes this approach computationally feasible is the fact that we are only attempting to recover the lighting parameters, while fixing the shape parameters. Otherwise, the problem would be massively underconstrained. This should also allow the use of more sophisticated lighting models [16], [36] to relax the single light and Lambertian assumptions.

4.2 Temporal Consistency

When video sequences are available, we can rely on temporal consistency between consecutive shapes to select the most likely ones. Let us assume that a second order autoregressive model [2] has been learned from training data. Given such a model, the shape at time t, $\mathbf{x}^t$, can be expressed as function of the shapes at times $t-1$ and $t-2$ as

$$\mathbf{x}^t = \hat{\mathbf{A}}_2 \mathbf{x}^{t-2} + \hat{\mathbf{A}}_1 \mathbf{x}^{t-1} + \hat{\mathbf{B}} \mathbf{w}^t \ , \tag{20}$$

where $\hat{\mathbf{A}}_2$, $\hat{\mathbf{A}}_1$ and $\hat{\mathbf{B}}$ are $3n_v \times 3n_v$ matrices learned offline, and $\mathbf{w}^t$ is an n_v Gaussian noise vector.

For any three consecutive images and the corresponding shape samples, the most plausible shape in the third one can be found by considering all $\{\mathbf{x}_i^{t-2}, \mathbf{x}_j^{t-1}, \mathbf{x}_k^t\}$ triplets and picking the $\mathbf{x}_k^t$ belonging to the one that best satisfies Eq. 20. Since this is done independently at each time step t, we are not imposing temporal consistency beyond our three consecutive frames windows.

5 Results

In this section we will evaluate the performance of the proposed algorithm. Since one of the core elements of our approach is the exploration strategy described in Sect. 3, we will first perform an analysis that will bring to light its benefits compared to alternative strategies for exploring the solution space. We will then evaluate the whole methodology for recovering non-rigid shape, and will compare to other state-of-the art approaches.

5.1 Analysis of the Exploration Strategy

The exploration strategy proposed in Sect. 3.2, accumulates samples which are drawn in several batches from a multivariate Gaussian. The success of the methodology depends on two key ingredients:

1) Initialization of the distribution grounded on the propagation of uncertainty from the image plane to the shape space, as described in Section 3.2.1.

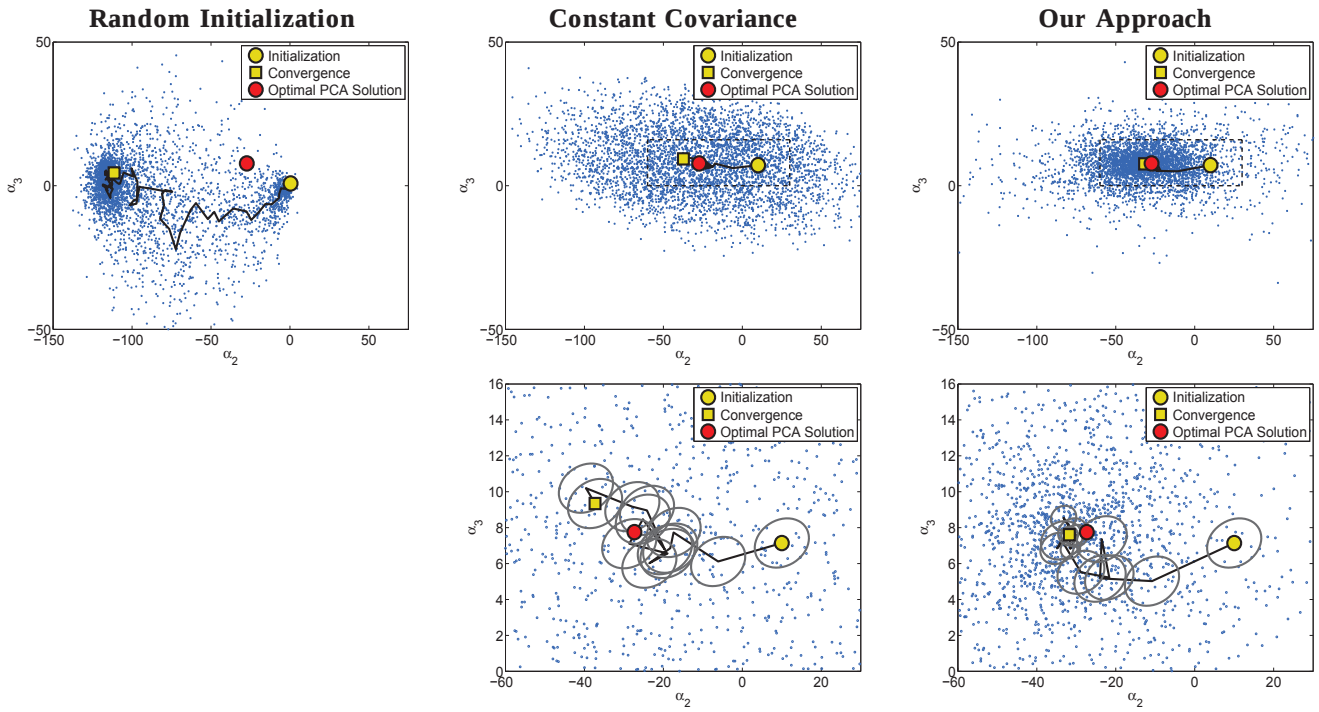

Fig. 7. Comparison of three sampling strategies to explore the solution space. **Upper Row:** View of the path followed during the exploration and the distribution of the accumulated samples. Note in the top-left graph that a random initialization of the search converges far from the optimal solution, and that the number of accurate samples is negligible. **Bottom Row:** Close up comparison of our approach, and a method in which the covariance of the distribution is not updated. Although both methods converge on a similar solution, the adaptation of the covariance matrix yields a higher concentration of samples close to the optimal solution.

2) Adapting both the mean and the covariance of the distribution as on the CMA algorithm [14], as discussed in Section 3.2.2.

We will now show the importance of these two design choices. To this end, we will compare *Our Approach* to two similar ones:

- *Random Initialization:* The initial distribution is chosen to be isotropic, with a random mean and a covariance taken to be the identity matrix. Both the mean and covariance are updated using the CMA approach.
- *Constant Covariance:* The mean and covariance are initialized according to Eq. 7 and Eq. 8, respectively. During the exploration process, only the mean is updated. The covariance is kept constant.

To perform the analysis we built a synthetic data set by deforming an initially planar 9×9 mesh of 30×30 cm. We created 400 meshes such as the one of Fig. 1 by randomly changing the angles between neighboring facets. In addition, we computed the deformation modes by applying Principal Component Analysis over a distinct set of 100 meshes generated in a similar manner. We then placed a virtual camera approximately 75 cm above the mesh and produced 100 random 3D-to-2D correspondences between the reference configuration and each of the individual deformed meshes. Finally, a 2-pixel standard deviation Gaussian noise was added to the 2D coordinates.

Given the deformation modes and the set of 3D-to-2D correspondences we then explored the modal weight space with each of the aforementioned strategies, and computed the statistics of the accumulated samples. In particular, we computed the distribution of the reprojection and inextensibility errors as defined in Eqs. 13 and 14, and the distribution of the reconstruction error defined by:

$$\texttt{Rec_Err}(\tilde{\boldsymbol{\alpha}}) = \sum_{i}^{n_v} \|\tilde{\mathbf{v}}_i - \mathbf{v}_i\| \,, \tag{21}$$

where $\mathbf{x}^{\texttt{true}} = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_{n_v}^\top]^\top$ corresponds to the ground truth shape we seek to recover and $\tilde{\mathbf{x}} = [\tilde{\mathbf{v}}_1^\top, \ldots, \tilde{\mathbf{v}}_{n_v}^\top]^\top$ is the shape corresponding to a sample $\tilde{\boldsymbol{\alpha}}$.

Fig. 6 depicts the average results over all the 400 deformed meshes. The solid lines show the error distribution of the samples accumulated by each method, and confirm that an approach as the one we propose in which the covariance matrix is iteratively adapted yields a larger concentration of samples with small reconstruction error than an approach in which the covariance is kept constant. From these curves, we can observe that a random initialization of the method does not guarantee a sufficient number of accurate samples and converges very far away from the optimal solution.

The dashed vertical lines correspond to the error of the mean $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^k$ in Eq. 15 at the convergence of the exploration strategies. Note that this is the solution that

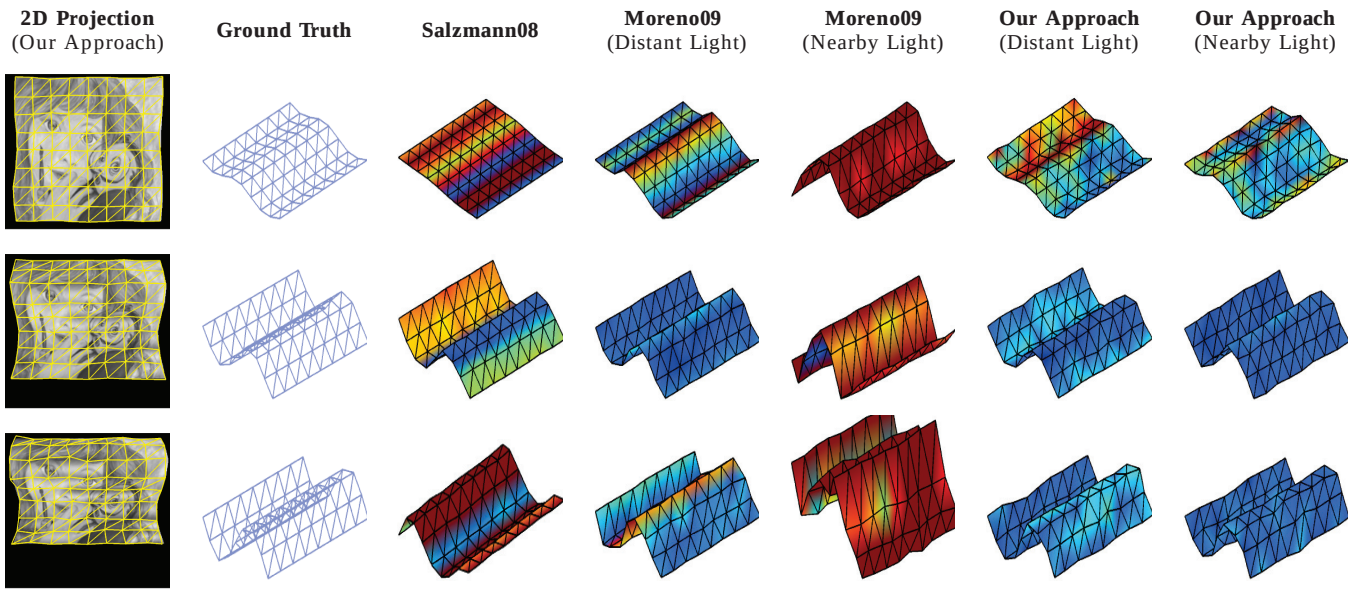

Fig. 8. Reconstruction results for the synthetic wave sequence. They are best viewed in color as deviations from the ground truth are encoded according the color-code of Fig. 3. Errors of more than 75% of the maximum amplitude of the ground truth shape appear in red.

would be taken if only reconstruction and inextensibility errors were considered. However, as we have argued throughout the paper, this does not guarantee minimizing the reconstruction error. In our approach we handle this issue by retaining several potential solutions which are subsequently evaluated under more discriminative constraints. In fact, as seen from the intersection of the dashed and solid lines, approximately a 70% of the samples in our approach are more accurate than the shape estimated from the mean of the distribution.

To provide a reference of the magnitude of the error, we also plot the results of an *Optimal PCA Solution* $\boldsymbol{\alpha}^{\mathtt{opt}}$, computed by projecting the ground truth shape $\mathbf{x}^{\mathtt{true}}$ onto the deformation modes. From Eq.1,

$$\boldsymbol{\alpha}^{\mathtt{opt}} = (\mathbf{Q}^\top \mathbf{Q})^{-1}\mathbf{Q}^\top(\mathbf{x}^{\mathtt{true}} - \mathbf{x}_0)\ . \tag{22}$$

This solution corresponds to the best reconstruction that may be obtained when approximating the ground truth shape by a linear combination of deformation modes. Note again in Fig.6, that a small reconstruction error is not directly correlated with small inextensibility and reprojection errors.

Finally, Fig. 7 shows one, but representative example of how the samples are distributed for each of the techniques compared in this section. Observe that our strategy clearly concentrates a larger amount of samples close to the optimal solution than other methods.

5.2 Non-Rigid Shape Recovery

For evaluating the complete algorithm, we compare its performance on two synthetic and two real image sequences against that of two state-of-the-art techniques [28], [23], which we refer to as *Salzmann08* and *Moreno09*, respectively. As discussed in Section 2, the first essentially returns the approximate solution of Eq. 6 that minimizes the variations in edge-length from the reference shape while the second returns the solution that best fits a shading model involving a point light source at infinity.

5.2.1 Synthetic Results

Besides the synthetic random meshes described in the previous section, we built another data set of 250 meshes by giving the initially planar 9×9 mesh of 30×30 cm a wave-like shape, as shown in Fig. 8. In both experiments, we used a real image as a texture-map and synthesized shaded images by selecting a random light-source direction in the hemisphere above the mesh. The light was located either infinitely far or within 30 cm of the mesh center. To compare the sensitivity of Moreno09 and of our approach to lighting conditions, for each synthetic shape we computed two different estimates, one using the image rendered using the distant light source and the other using the nearby light.

Fig. 8 depicts the reconstruction results on three frames of the synthetic wave sequence using Salzmann08, Moreno09, and our own approach in conjunction with either the distant or the nearby lighting.

In the first two rows of Fig. 9, we use boxplots[1] to quantitatively summarize all synthetic results. We plot the mean reconstruction, reprojection and inextensibility errors, as defined by Eqs. 21, 13 and 14, respectively. In addition, besides computing the error of Salzmann08, Moreno09, Our Approach and the Optimal PCA Solution, we also include the output of an hypothetical

1. Box denoting the first $\mathcal{Q}1$ and third $\mathcal{Q}3$ quartiles, a vertical line indicating the median, and a dashed line representing the data extent taken to be $\mathcal{Q}3+1.5(\mathcal{Q}3-\mathcal{Q}1)$. The crosses denote points lying outside of this range.

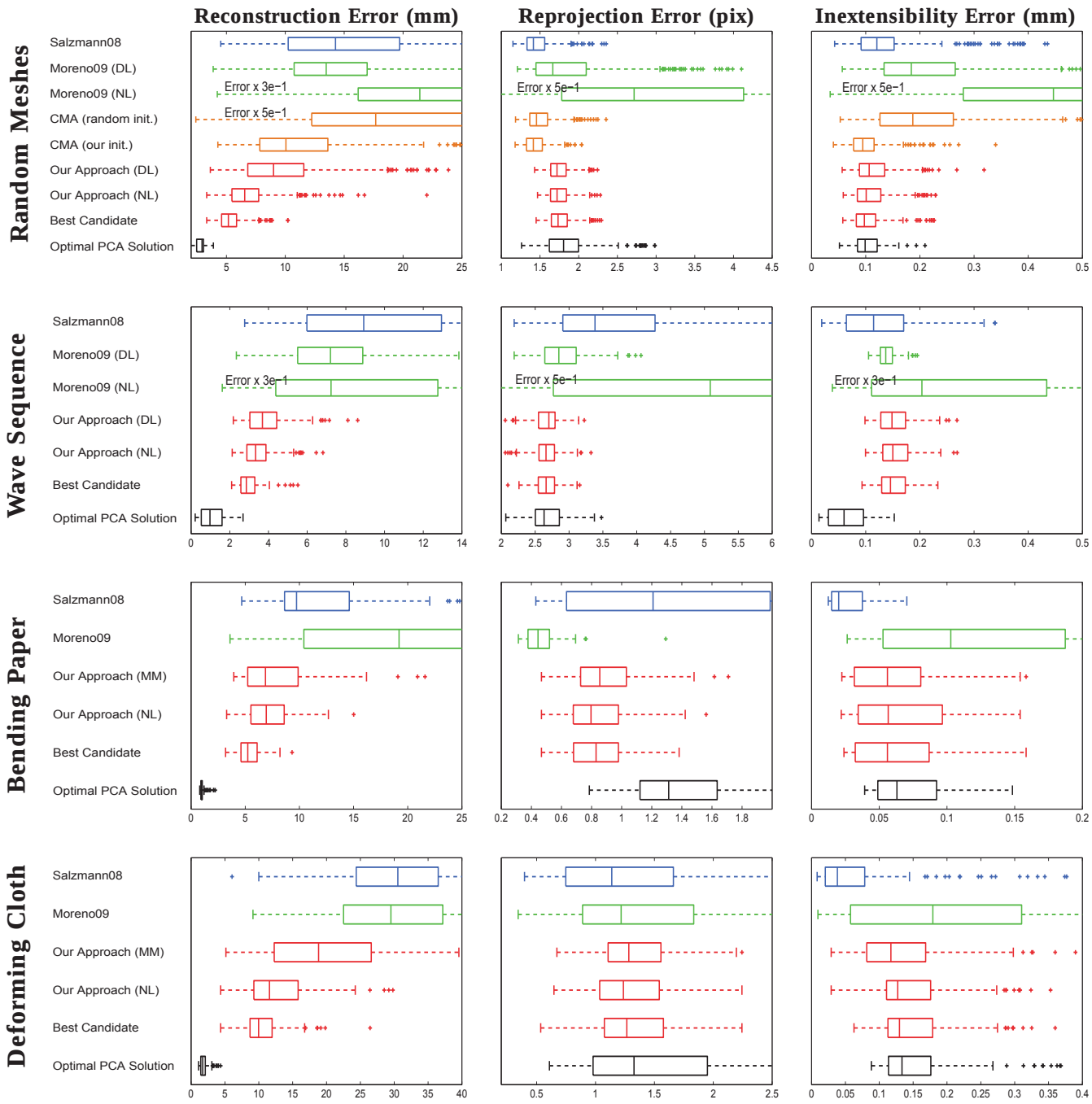

Fig. 9. In each row, reconstruction, reprojection, and inextensibility errors for each of the two synthetic and the two real sequences. DL: Distant Light. NL: Nearby Light. MM: Motion Model. Note that some of the errors are scaled to fit within the figure.

algorithm that would be able to select the *Best Candidate* shape among all the samples produced by the sampling mechanism of Sect. 3, which represents the theoretical optimum an algorithm like ours could achieve by using the image information as effectively as possible. Furthermore, for the experiment with random meshes we plot the errors of the CMA algorithm [14] when is not used within our exploration framework, that is, when it is directly used to minimize the objective function of Eq. 12. We consider both the case when CMA is randomly initialized, and the case when it is initialized to the Gaussian distribution in the modal weight space we propose in Sect. 3.2.1.

From all these errors, we can observe that our method consistently returns a lower 3D reconstruction error, which shows that it is more accurate than the other methods and very close to the optimum. This is true even though the reprojection and inextensibility errors are very similar for all the methods, which confirms that minimizing these is not sufficient by itself to retrieve the

	Sa	Mo		OA			BC
		DL	NL	DL	NL	MM	
Random Meshes	84	81	15	91	99	–	100
Wave Sequence	78	95	31	100	100	–	100
Paper Bending	80	–	43	–	97	96	100
Deforming Cloth	59	–	57	–	97	81	99

TABLE 2
Percentages of correct solutions for all four set of experiments with non-rigid surfaces. Sa: Salzmann08. Mo: Moreno09. OA: Our Approach. BC: Best Candidate. DL: Distant Light. NL: Nearby Light. MM: Motion Model.

correct 3D shape.

Both Moreno09 and Our Approach address this issue by taking advantage of shading cues. Since we explicitly model a nearby light, we clearly outperform Moreno09 in that case. Less intuitively, we also outperform it in the distant light case, even though we then use the same simple shading model. We believe this is due to the fact that we use both the inextensibility and shading constraints, whereas Moreno09 uses only the latter.

Another measure of success is the *Percentage of correct solutions* of Table 2. Given the ground truth solution, a 3D sample mesh is considered to be correct if at least 75% of its vertices have a reconstruction error smaller than $0.5\times$Height, where *Height* refers to the maximum amplitude of the ground truth shape. Again, our approach clearly yields the best numbers. The specific ratios –75% and $0.5\times$Height– are of course *ad hoc* and have been chosen so that 3D meshes that are deemed incorrect produce disturbing effects when viewed in sequence. To provide the reader with an intuitive understanding of what this measure actually represents, in Fig. 8 facets with reconstruction errors of more than 75% are color-coded in red.

Finally, the table at the top of Fig. 10 depicts the accuracy of the estimated lighting parameters. Note that we estimate the position and direction of a light source that was allowed to move freely within a 30 cm radius hemisphere with an error below 1 cm and 10 degrees, respectively.

5.2.2 Real Images

We also tested our approach on a 120-frames sequence of a bending paper and a 150-frame sequence of a deforming T-shirt, both acquired with a Pointgrey Bumblebee stereo rig. The surfaces were lit by a dim ambient lighting and a light source located at about 30 cm from the surface.

We used the stereo pairs to estimate the ground truth shape and then ran our algorithms using the output of a single camera. We used SIFT [18] to establish correspondences between the reference and input images. Erroneous ones were removed by initially fitting a deformable smooth 2D mesh and discarding inconsistent matches, as in [23].

In both cases we used the algorithm described in Section 3 to initially produce a set of candidate 3D shapes

	Distant Light	
	Direction Error (deg)	Power Error (%)
Random Meshes	6.9 ± 4.3	5.2 ± 2.1
Wave Sequence	2.1 ± 0.9	2.2 ± 0.8
	Nearby Light	
	Position Error (mm)	Power Error (%)
Random Meshes	7.4 ± 6.1	6.8 ± 3.3
Wave Sequence	3.2 ± 0.8	2.8 ± 1.0

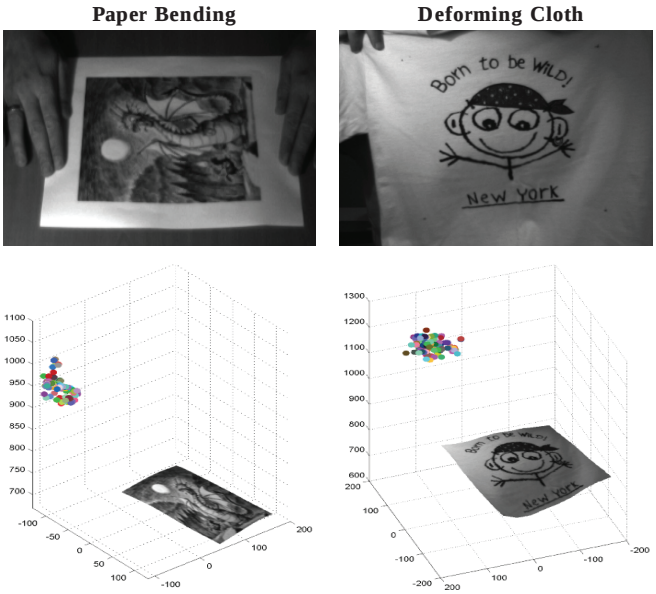

Fig. 10. Estimated lighting parameters. Upper tables: Mean error and standard deviation of the lighting parameters –direction, position and power– estimated independently in each frame of the synthetic sequences. Bottom figures: Light source positions estimated independently in all frames of the real sequences.

in each individual frame. We then chose the best using either shading or motion information.

When using shading we assumed a situation of a nearby light source. The reconstruction errors depicted in the boxplots of Figure 9 exhibit the same patterns as those obtained for the synthetic sequences, which confirms that our method outperforms the other two. As shown in Table 2, we obtain 97% of correct solutions, which represents a 30% increase in performance, using the same definition of "correct" as before. In the bottom of Figure 10, we plot the estimated light source positions in each frame. Although we did not measure the exact light source locations, the fact that the estimates are tightly clustered is an indication that they are probably correct, given that they all were obtained independently.

Finally, we also evaluated our approach when using motion to disambiguate. To learn the autoregressive model of Section 4.2, we acquired additional sequences, obtained ground truth data using our stereo rig, and learned the model parameters by probabilistic fitting [2]. In the case of the sheet of paper, as shown in the third row of Figure 9 and in Table 2, using the motion model yields similar results to those obtained using

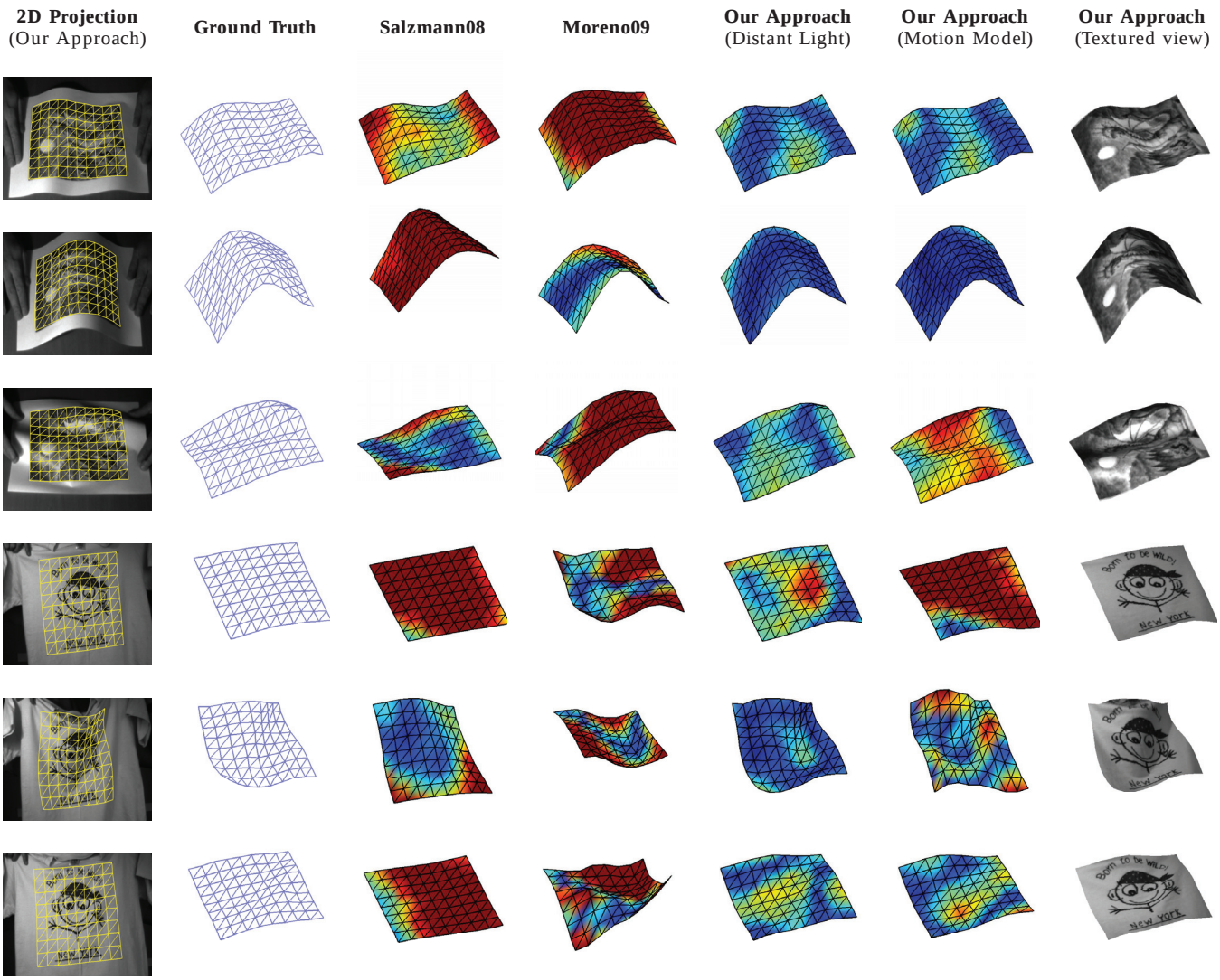

Fig. 11. Reconstruction results for the two real sequences. Top three rows: Paper Bending. Bottom three rows: Deforming Cloth. The reconstruction errors are again color-coded.

shading and clearly outperforms both Salzmann08 and Moreno09. The performance degrades slightly in the case of cloth because our second order motion model is not accurate enough to perfectly capture the dynamics of the sharp cloth deformations. Nevertheless, our algorithm still outperforms the other two state-of-the-art methods.

6 Conclusion

Geometry-based approaches to single view 3D reconstruction can easily return erroneous solutions that both satisfy the constraints and yield plausible reprojections. To overcome this problem, given that the input data is noisy, we use error propagation techniques to derive an analytical expression of the space of potential solutions and to propose a small but representative number of samples. The best among them can then be chosen using additional image cues, such as shading or motion, which significantly improves results at a limited computational cost.

We have demonstrated the effectiveness of this approach for 3D deformable shape reconstruction. However, our approach relies on representing deformation as weighted sums of deformation modes, which somewhat reduces its applicability. In future work, we will therefore seek to extend it using first order matrix perturbation theory [9] to reason directly in the space of vertex positions. While applying this relaxation of the problem would require having to consider much larger sampling densities, we believe it can be feasibly addressed in the context of specialized Monte Carlo techniques such as the Partitioned Sampling proposed in [19].

Acknowledgments

We would like to thank Josep M Porta and Mathieu Salzmann for insightful comments and suggestions on a preliminary version of this work. We would also like to thank Aaron Hertzmann for bringing the CMA algorithm to our attention.

REFERENCES

[1] A. Balan, M. Black, H. Haussecker, and L. Sigal. Shining a light on human pose: On shadows, shading and the estimation of pose and shape. In *International Conference on Computer Vision*, pages 1–8, 2007.

[2] A. Blake and M. Isard. *Active contours*. Springer, 1998.

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. ACM SIGGRAPH*, pages 187–194, 1999.

[4] M. Brand. Morphable 3D models from video. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

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

[6] A. Chiuso, R. Brockett, and S. Soatto. Optimal structure from motion: Local ambiguities and global estimates. *International Journal of Computer Vision*, 39(3):195–228, 2000.

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

[8] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *European Conference on Computer Vision*, pages 484–498, 1998.

[9] A. Criminisi. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. Springer, 2001.

[10] M. de La Gorce, N. Paragios, and D. Fleet. Model-based hand tracking with texture, shading and self-occlusions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[11] A. Ecker, A. D. Jepson, and K. N. Kutulakos. Semidefinite programming heuristics for surface reconstruction ambiguities. In *European Conference on Computer Vision*, pages 127–14, 2008.

[12] N. Gumerov, A. Zandifar, R. Duraiswami, and L. Davis. Structure of applicable surfaces from single views. In *European Conference on Computer Vision*, pages 482–496, 2004.

[13] G. Hamerly and C. Elkan. Learning the K in K-Means. In *Neural Information Processing Systems*, 2003.

[14] N. Hansen. The CMA evolution strategy: a comparing review. In *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, pages 75–102. Springer, 2006.

[15] N. Hansen. The CMA evolution strategy: A tutorial. *Available Online at: http://www.bionik.tu-berlin.de/user/niko/cmatutorial.pdf*, 2007.

[16] K. Hara, K. Nishino, and K. Ikeuchi. Light source position and reflectance estimation from a single view without the distant illumination assumption. *IEEE Transactions Pattern Analylis and Machine Intelligence*, 27(4):493–505, 2005.

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

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

[19] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *European Conference on Computer Vision*, pages 3–19, 2000.

[20] 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, 1993.

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

[22] F. Moreno-Noguer, J.M. Porta, and P. Fua. Exploring ambiguities for monocular non-rigid shape estimation. In *European Conference on Computer Vision*, pages 361–374, 2010.

[23] F. Moreno-Noguer, M. Salzmann, V. Lepetit, and P. Fua. Capturing 3D stretchable surfaces from single images in closed form. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1842–1849, 2009.

[24] M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito. Optimal metric projections for deformable and articulated structure-from-motion. *International Journal of Computer Vision*, pages 1–25, 2011.

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

[26] G. Qian and R. Chellappa. Structure from motion using sequential montecarlo methods. *International Journal of Computer Vision*, 59(1):5–31, 2004.

[27] M. Salzmann and P. Fua. Linear local models for monocular reconstruction of deformable surfaces. *IEEE Transactions Pattern Analylis and Machine Intelligence*, 33:931–944, 2011.

[28] 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*, volume 4, pages 581–594, 2008.

[29] J. Sanchez, J. Ostlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1189–1196, 2010.

[30] R. Szeliski and S. Kang. Shape ambiguities in structure-from-motion. *IEEE Transactions Pattern Analylis and Machine Intelligence*, 5:506–512, 1997.

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

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

[33] Y. Wang, Z.C. Liu, G., Hua, Z. Wen, Z.Y. Zhang, and D. Samaras. Face re-lighting from a single image under harsh lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[34] R. White and D.A Forsyth. Combining cues: Shape from shading and texture. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1809–1816, 2006.

[35] J. Xiao and T. Kanade. Uncalibrated perspective reconstruction of deformable structures. In *International Conference on Computer Vision*, pages 1075–1082, 2005.

[36] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *ACM SIGGRAPH*, pages 215–224, 1999.

[37] J. Zhu, S. Hoi, Z. Xu, and M. Lyu. An effective approach to 3D deformable surface tracking. In *European Conference on Computer Vision*, pages 766–779, 2008.

Francesc Moreno-Noguer received the MSc degrees in industrial engineering and electronics from the Technical University of Catalonia (UPC) and the Universitat de Barcelona, in 2001 and 2002, respectively. He obtained his PhD degree from UPC in 2005, and his work received the UPC's Doctoral Dissertation Extraordinary Award. From 2006 to 2008 he was a postdoctoral fellow at the computer vision departments of Columbia University and the École Polytecnique Fédérale de Lausanne. In 2009 he joined the Institut de Robòtica i Informàtica Industrial in Barcelona, as an associate researcher of the Spanish Scientific Research Council. His research interests are focused on retrieving rigid and non-rigid shape, motion and camera pose from single images and video sequences.

Pascal Fua received an engineering degree from Ecole Polytechnique, Paris, in 1984 and the Ph.D. 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, analysis of microscopy images, and Augmented Reality. He has (co)authored over 200 publications in refereed journals and conferences. He is an IEEE Fellow and has been an associate editor of IEEE journal Transactions for Pattern Analysis and Machine Intelligence. He often serves as program committee member, area chair, and program chair of major vision conferences.