Multiview 3D Warps

Alessio Del Bue
Istituto Italiano di Tecnologia (IIT)
alessio.delbue@iit.it

Adrien Bartoli
ALCoV - ISIT, Université Clermont 1
adrien.bartoli@gmail.com

Abstract

Image registration and 3D reconstruction are fundamental computer vision and medical imaging problems. They are particularly challenging when the input data are images of a deforming body obtained by a single moving camera. We propose a new modelling framework, the multiview 3D warps. Existing models are twofold: they estimate inter-image warps which are often inconsistent between the different images and do not model the underlying 3D structure, or reconstruct just a sparse set of points. In contrast, our multiview 3D warps combine the advantages of both; they have an explicit 3D component and a set of 3D deformations combined with projection to 2D. They thus capture the dense deforming body's time-varying shape and camera pose. The advantages over the classical solutions are numerous: thanks to our feature-based estimation method for the multiview 3D warps, one can not only augment the original images but also retarget or clone the observed body's 3D deformations by changing the pose. Experimental results on simulated and real data are reported, confirming the advantages of our framework over existing methods.

1. Introduction

Image registration and visual 3D reconstruction have become major research topics over the last few decades, for they lie at the heart of a large body of applications. While (multimodal) medical image registration is an older field, monocular 3D reconstruction of deforming bodies has only recently been investigated in computer vision (see for instance [3, 16].) In both cases, the problem of relating a set of images showing the same body under different pose and deformation quickly became major in the field. Despite significant achievements, it is still an open problem. Many approaches are based on estimating so-called *warps*. Warps are image deformation functions that relate corresponding points across the input images. Among them, the Thin-Plate Spline (TPS) warp is well-known and extensively used [2].

Existing parametric image warps in the literature such as the TPS warp map point coordinates to point coordinates. Registering $n \geq 2$ images thus entails one to estimate $n-1$ warps at least, so that by 'chaining' warps, every pair of images in the image set can be related and compared. This approach has several drawbacks. First, many warps such as the TPS warp are not part of a group – they can thus not be easily 'chained' since warp composition and inversion are not properly defined [11]. Second, it is difficult with this model to use all the available image information at the warp estimation phase (if images 1 and 2 get registered, as well as images 2 and 3, the coupling between images 3 and 1 is ignored.) It is worth of note that some work address the particular problem of finding groupwise registration of an image set [6, 19] but do not use a 3D modelling.

In this paper, we propose a novel framework to the parametric modelling of image warps for $n \geq 2$ images. Our framework is fundamentally different from the literature in that it takes into account by construction the fact that multiple images must be modelled. In essence, an image results of a formation process (whether it is an optical image or an MRI slice for instance.) Our framework is generative: we propose the *multiview 3D warps* that model a set of images of a deforming body. Inspired by the geometric image formation process, it has two major components: *(i)* a 3D component acting as an abstraction of the body observed in the images and *(ii)* a set of 3D to 2D deformation functions that each combine a 3D deformation and a projection to 2D, thereby encapsulating the body's deformation and pose for every image. This provides a clear advantage over previous 2D warps which were encapsulating pose, camera projection and deformation in a single function [1]. Differently, by placing the warp directly in a metric 3D space, we detach the 3D deformations from the actual camera projection and pose variations. Thus, the learned warp may be reused in other imaging scenarios and with viewpoints different from the original ones. Despite the modelling details that must be solved to implement our framework, we tackle several other issues it raises: the estimation and the inter-image point transfer problems.

Figure 1 schematically summarises the multiview 3D warps and their features in a general image modelling scenario. The first problem is to, given a set of keypoints or

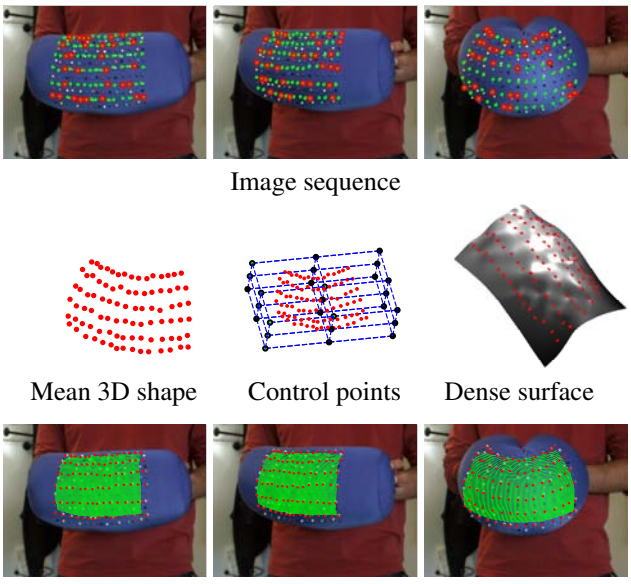


Figure 1. Our multiview 3d warps in a few figures. The first row shows an image sequence with green dots representing the image tracks and red circles the missing data. The second row shows the warping formation and augmentation in three stages: 1) From the point tracks we extract a mean 3D shape using SfM, 2) we place a set of control points around the 3D shape (black dots) and we learn the 3D warping functions given the image data, 3) we augment the 3D shape with a dense surface obtained by simple interpolation. The last row shows the reprojection of the dense surface warped to model the image deformations.

anatomical landmarks matched across multiple images, infer the parameters of our 3D warp (both the 3D component and the set of 3D to 2D deformation functions); the second problem is to be able to transfer points from one to another image of the set given the 3D warp parameters. Notice that, differently from the classical 2D case, the warp's control points are not manually placed in a reference frame but automatically computed in 3D space with a general purpose Structure-from-Motion (SfM) algorithm. We give general solutions to these problems. Our generic warp will be implemented in the framework of Radial Basis Functions (RBF). Notice that we use a multiquadric kernel but any other kernel could be used instead.

2. State of the Art

Most of the literature on image registration and monocular 3D reconstruction of a deforming body uses inter-image warps, that will map a point from one image to a point in another image. Given a set of $n \geq 2$ images, existing registration approaches simply estimate $n - 1$ warps.

The literature on inter-image warps is dense. Based on Duchon's Thin-Plate Spline (TPS) [9], Bookstein proposed the famous TPS warp [2] for the registration of anatomical landmarks. It was later extended in a number of ways. A work related to ours is the Generalized TPS (GTPS) warps of Bartoli *et al.* [1]. While it is shown that the TPS warp implements the affine projection of some deforming surface, 3D warps were derived that implement the perspective projection of rigid and deforming surfaces between two images. Other popular warps found in the literature are Free-Form Deformations (FFD) warps. They use the tensor product of two 1D smooth functions, typically the cubic B-spline [15]. A 3D modelling based on the FFD has recently been shown to lead to the so-called NURBS warp [4]. It is also possible to model a warp by a triangular mesh; this is a common choice in deformable surface tracking [14] and 3D reconstruction [16]. The approach of [16] differs from ours mainly in that it uses a model whereby a 3D surface is deformed and projected to the images into two distinct steps, whereas we aggregate those two steps into a single one.

Finally, there exist work in medical image processing on groupwise image registration that addresses the multi-view problem, but since the medical data are fundamentally different from the data in computer vision, they do not use a full 3D modelling of the imaged body [6, 19]. The literature thus contains some works on 3D warps, but limited to image pairs, or multi-image warp models, but which are not based on a real 3D modelling. The major difference with our work is that we do not split the image set to be modeled in image pairs, but rather estimate a single 3D warp that covers the relationship between all images in the set.

Another stream of research aims to reconstruct sparse 3D point sets of deformable objects. This is the Non-Rigid Structure-from-Motion problem [3] where the camera projection matrices and the underlying deformations of the shape are learned solely from 2D image trajectories. They are based on the Low-Rank Shape Model (LRSM) that represents the 3D body shape as a combination of deformation modes. The literature is vast and most recent methods provides reasonable results for smooth deformations [18, 7, 13, 8]. More recently, approaches considering objects as piecewise rigid deal with a wider range of deformations [17, 10]. Differently from warps, the deformation modelling is restricted to a sparse set of points and does not directly generalize to surfaces. In our framework, we recover a dense mapping (and not just sparse 3D points); the LRSM can be used to constrain the 3D recovered shape.

3. The Multiview 3D Warp

The gist of the approach is to consider the warping function embedded directly in the metric 3D space and not onto the image plane like classical image warps. Learning such a warp solely from the image data results in an ill-posed problem especially if deformations are involved, therefore

priors on the 3D warp are required. In the following, we first derive our multiview 3D warp model and then review several priors that are used within our framework.

3.1. Derivation

For the sake of simplicity, we have chosen to implement our general warp using an RBF. The warp is driven by a set of 3D centres $\mathbf{c}_k$ with $k = 1 \dots l$. A 3D point $\mathbf{x}$ is mapped to its projection in one of the images by undergoing a 3D transform followed by a projection. We assume for now that the warp's parameterisation – the centres and 3D points for each observed image point – is known; we provide a way to estimate these parameters from the image data in Sec. 4.2. The 3D transform is applied for forming the following vector that contains the distance of point $\mathbf{x}$ to the centres:

$$\mathbf{l}_{\mathbf{x}} = \begin{bmatrix} \rho\left(d^2(\mathbf{x}, \mathbf{c}_1)\right) \\ \vdots \\ \rho\left(d^2(\mathbf{x}, \mathbf{c}_l)\right) \\ \mathbf{x} \end{bmatrix}, \tag{1}$$

where ρ is the kernel function of a given distance measure. We use the Euclidean distance and a multiquadric RBF: $\rho(d) = \sqrt{d + \beta}$ where $\beta \in \Re$. Given a set of 3D points $\mathbf{x}_j$ with $j = 1 \dots p$ we can form the $p \times (l + 3)$ *warp transfer matrix* $\mathtt{L}$ which contains all the point-to-centre distances and take the l 3D centres in an $l \times 3$ matrix $\mathtt{P}$:

$$\mathtt{L} = \begin{bmatrix} \mathbf{l}_1^\top \\ \vdots \\ \mathbf{l}_p^\top \end{bmatrix} \text{ and } \mathtt{P} = \begin{bmatrix} \mathbf{c}_1^\top \\ \vdots \\ \mathbf{c}_l^\top \end{bmatrix}. \tag{2}$$

Following the standard feature-based estimation of an RBF (see for instance [2, 1]) we apply the warp to the l centres and form matrix $\mathtt{K}_\lambda$ (with λ some small internal smoothing coefficient) as:

$$\mathtt{K}_{m,n} = \begin{cases} \lambda & m = n \\ \rho\left(d^2(\mathbf{c}_m, \mathbf{c}_n)\right) & \text{otherwise.} \end{cases} \tag{3}$$

Solving for the warp coefficients is then done in closed-form using matrix $\mathtt{E}_\lambda$ of size $(l + 3) \times l$:

$$\mathtt{E}_\lambda = \begin{pmatrix} \mathtt{K}_\lambda^{-1}(\mathtt{I} - \mathtt{P}(\mathtt{P}^\top \mathtt{K}_\lambda^{-1} \mathtt{P})^{-1} \mathtt{P}^\top \mathtt{K}_\lambda^{-1}) \\ (\mathtt{P}^\top \mathtt{K}_\lambda^{-1} \mathtt{P})^{-1} \mathtt{P}^\top \mathtt{K}_\lambda^{-1} \end{pmatrix}. \tag{4}$$

Matrix $\mathtt{E}_\lambda$ maps the warp centres' coordinates to the 'natural' RBF parameters. Given a set of f images and a set of p points matched between images, we can write the optimisation problem for the 3D warp at frame i as:

$$\min_{\mathtt{P}_i, \mathtt{M}_i} \left\| \mathtt{L}\mathtt{E}_\lambda \mathtt{P}_i \mathtt{M}_i - \mathtt{Q}_i \right\|^2, \tag{5}$$

where $\mathtt{Q}_i$ is a $p \times 2$ matrix containing the image points and $\mathtt{M}_i$ is the 3×2 affine camera matrix at frame i. In this work we consider the simplest orthographic camera model (i.e. $\mathtt{M}_i^\top \mathtt{M}_i = \mathtt{I}_2$ where $\mathtt{I}_2$ is a 2×2 identity matrix.) Notice that the image coordinates stored in $\mathtt{Q}_i$ are registered to the image centroid at each frame i.

To obtain the most compact formulation, we can stack together the varying 2D coordinates at each frame:

$$\mathtt{L}\mathtt{E}_\lambda \underbrace{\begin{bmatrix} \mathtt{P}_1 & \dots & \mathtt{P}_f \end{bmatrix}}_{\mathtt{P}} \underbrace{\begin{bmatrix} \mathtt{M}_1 & \dots & \mathtt{0} \\ \vdots & \ddots & \vdots \\ \mathtt{0} & \dots & \mathtt{M}_f \end{bmatrix}}_{\mathtt{M}} = \underbrace{\begin{bmatrix} \mathtt{Q}_1 & \dots & \mathtt{Q}_f \end{bmatrix}}_{\mathtt{Q}} \tag{6}$$

Notice here that matrices $\mathtt{L}$ and $\mathtt{E}_\lambda$ are fixed at each frame. Deformations and shape motion are instead described by the time-varying components in the overall camera matrix $\mathtt{M}$ and the varying control points $\mathtt{P}$. It is also important to stress that we do not have an explicit Euclidean localisation of the deforming 3D points in time. The deformation of the control points $\mathtt{P}$ solely describe the 2D image deformations of the object. This is in contrast with previous approaches in SfM where the 3D deforming body was estimated explicitly. Differently with our 3D warps we define a deformation field by estimating a specific warping function at each frame. Moreover, this deformation field as defined in Equation (6) models the whole multi-view warping functions in contrast to previous approaches evaluating pairwise transformations between image frames.

3.2. Multiview 3D Warp priors

Equation (5) is under-constrained: there are more than one deformation that minimize the reprojection error. Choosing an arbitrary solution would not allow the multiview 3D warp to accurately capture the 3D body's average shape and deformations, and camera pose. So as to select a sensible solution, it is clear that priors must be introduced. Below, we review a set of priors that can be easily used within our framework. They are grouped into two classes: *physical priors* and *statistical priors*. The priors are expressed as penalties to be used when estimating the warp.

Physical priors. We identified three types of physical priors for our multiview 3D warps:

• The 3D deformation cannot deform the initialised control points position $\mathtt{P}_0$ arbitrarily. We can thus penalize the deviation from the initial configuration using $\chi_i^c = \|\mathtt{P}_0 - \mathtt{P}_i\|^2$.

• The warp's deformation energy is likely to be low.[1] It is expressed by $\chi_i^q = \text{trace}(\mathtt{P}_i^\top \bar{\mathtt{K}}_\lambda \mathtt{P}_i) = \|\mathtt{Z}\mathtt{P}_i\|^2$ where matrix $\bar{\mathtt{K}}_\lambda$ is given by removing the last three rows of $\mathtt{E}_\lambda$ and $\mathtt{Z}$ is any square root of $\mathtt{E}_\lambda$ obtained using Cholesky factorization, for instance.

[1] This deformation energy depends on the warps' implementation. For instance, in the case of a TPS, it is the integral bending energy.

• For video data, temporal smoothness can be favored using $\chi_i^t = \|\mathtt{P}_{i+i} - \mathtt{P}_i\|^2$.

Statistical priors. The warp deforms the body's shape but not in an unconstrained manner. The assumption that a deformation may be described by a set of deformation bases can be used in order to constrain the degrees of freedom of the warp. Such a prior has been successfully used in the Non-Rigid Structure-from-Motion (NRSfM) framework [3] to obtain a compact multi-view formulation of 3D deformations. Readapting this notion to our problem, we can constrain the body's 3D deformation using a set of *warp deformation bases* $\mathtt{B}_d$ which identifies D modes of variations. Each 3D control point configuration is then expressed as:

$$\mathtt{P}_i = \sum_{d=1}^{D} r_{id}\mathtt{B}_d \quad \text{with} \quad \mathtt{B}_d \in \Re^{l\times 3}, r_{id} \in \Re \tag{7}$$

where $r_{id} \in \Re$ are scalar weights that perform a linear combination of the basis $\mathtt{B}_d$. As opposed to the three physical priors, this statistical prior introduces several latent variables. Details on how we use these priors are given in Sec. 4.3 where we selected the mentioned statistical prior as the more suitable one for our modelling problem.

4. Estimation of the Multiview 3D Warp

We present a computational algorithm for estimating the 3D warps solely from a set of feature points extracted from a generic image sequence. We do not assume having complete matches for each image (the matrix of 2D measurements $\mathtt{Q}$ may have some missing entries.) Schematically we can summarise the algorithm in three steps:

1. Initial 3D feature points computation. Given 2D image matches, estimate a *mean* position of the 3D points $\mathbf{x}_j$ in metric space by running standard rigid SfM approach with missing data such as [12].

2. 3D warp placement. Given the mean position $\mathbf{x}_j$ find a bounding box enclosing the shape and evenly place the 3D control points $\mathbf{c}_k$. Compute $\mathtt{L}$ and $\mathtt{E}_\lambda$ given the $\mathbf{x}_j$ and the location of the warp's control points.

3. 3D warp optimisation. Optimise the multiview cost in Equation (5) given the control points and priors.

4.1. Initial 3D feature points computation

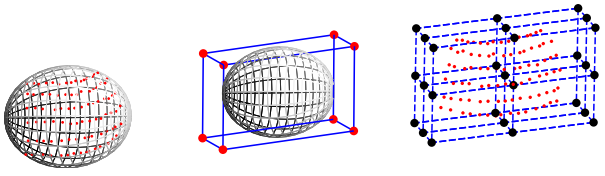
This stage is necessary to define the position of the control points and the warp kernel given a mean shape. Starting from a set of 2D matches in the images, we collect 2D image trajectories in the matrix $\mathtt{Q}$. Notice that missing point are a common occurrence, especially when the shape is deforming. Thus, we define the $2f \times p$ mask matrix $\mathtt{D}$ defining a known entry with a 1 and the missing coordinate with a 0. The optimisation problem is the following:[2]

$$\min_{\mathtt{R},\mathtt{S},\mathbf{t}} \left\| \mathtt{D} \odot (\mathtt{Q}^\top - \mathtt{R}\,\mathtt{S} + \mathbf{t}\,\mathbf{1}_p^\top) \right\|^2 \tag{8}$$

where $\mathtt{R}$ is a $2f \times 3$ matrix containing the orthographic camera matrices, $\mathtt{S}$ is a $3 \times p$ matrix containing the 3D metric coordinates for each feature point $\mathbf{x}_j$ (i.e. $\mathtt{S} = [\mathbf{x}_1, \ldots, \mathbf{x}_p]$), $\mathbf{t}$ is a $2f$-vector of the 2D image centroid for the set of points and $\mathbf{1}_p$ is a vector of p ones. This problem is classical SfM with missing data; it can be solved with several approaches [5]. The method of Marques and Costeira [12] obtains the highest resilience to missing data; this is the method we have chosen in our implementation.

4.2. 3D warp initialisation

Given the set of mean 3D points, we can find an initial placement for the control points. We first find an approximate convex hull by fitting an ellipsoid to the cloud of 3D points. This initial envelope is used as a first guess for the whole shape reconstructed by the previous SfM approach. A volumetric grid is then placed to contain the ellipse and each edge of the grid is sampled with a fixed number of control points. For the minimal configuration we have $l = 8$ points, for most of the experimental tests a number of $l = 27$ control points were sufficient. Thus, given this procedure we obtain an initial configuration stored in a matrix $\mathtt{P}_0$ of size $l \times 3$. Other sampling strategies may be adopted but results were satisfactory with this regular sampling. Figure 2 shows a graphical example using the data shown in Figure 1. Given the location of the control points



a) b) c)

Figure 2. The red points represent the mean 3D shape extracted from a rigid SfM algorithm. a) we compute the minimum volume ellipse containing these points. b) it is then possible to fit efficiently a bounding box containing the ellipsoid. c) control points are inserted by evenly sampling the volume of the box.

$\mathbf{c}_k$, we can construct the matrix $\mathtt{L}$ and $\mathtt{E}_\lambda$ as in Eqs. (2) and (4) respectively.

4.3. 3D warp optimisation

After initialising the warp, we last need to find the displacement of the control points given the 2D deformations

[2] The symbol $\odot$ denotes the element-wise Hadamard matrix product.

and the camera matrices which relate the 3D warping function to the image plane. From Eq. (6) we have that $\mathtt{L}\mathtt{E}_\lambda\mathtt{P}\mathtt{M} = \mathtt{Q}$ with the bilinear factor $\mathtt{G} = \mathtt{P}\mathtt{M}$ denoting the projection of the 3D warp to the image plane. The *warp projection matrix* $\mathtt{G}$ of size $l \times 2f$ can be estimated by solving:

$$\min_{\mathtt{G}} \left\| \mathtt{Q}^\top - \mathtt{L}\,\mathtt{E}_\lambda\,\mathtt{G} \right\|^2 + \|p(\mathtt{G})\|^2 \tag{9}$$

where the term $\|p(\mathtt{G})\|^2$ represents additional linear prior terms as presented in Section 3.2. This regularized problem can be solved with standard linear least squares by adding the additional quadratic cost given by the priors.

After this first step, we have to enforce the fact that the *warp projection matrix* $\mathtt{G}$ is a bilinear factor of the 3D deforming control points $\mathtt{P}$ and the camera projection matrix $\mathtt{M}_i$. Now, in order to factorise $\mathtt{G}$ in the two components, we have to satisfy the non-linear constraints arising from the specific camera model. In this work we assume the simplest orthographic camera model (i.e. $\mathtt{M}_i^\top\mathtt{M}_i = \mathtt{I}_2$) but other types of camera can be easily adapted to this framework. The problem becomes the optimisation of the cost function:

$$\min_{\mathtt{M},\mathtt{P}} \|\mathtt{G} - \mathtt{P}\mathtt{M}\|^2 \quad \text{subject to} \quad \mathtt{M}_i^\top\mathtt{M}_i = \mathtt{I}_2 \tag{10}$$

Generally, in the presence of smooth deformations, the matrix $\mathtt{G}$ is rank constrained i.e. $rank(\mathtt{G}) \ll max\{2f, p\}$ thus we need again a set of priors in order to solve such an ill-posed problem. The most pertinent prior for such a problem is the statistical prior of considering the control points modelled by a set of basis shapes.

The optimisation using a set of basis shapes can be solved with one of the iterative solvers in the literature such as the BALM [8]. As an initialisation for the camera parameters we used the previously computed motion matrices $\mathtt{R}$ for the 3D control points placement (see Sec. 4.1.). We also fix the first basis $\mathtt{B}_1$ in Eq. (7) to be equal to $\mathtt{P}_0$. In such way we impose the time-varying control point positions $\mathtt{P}_i$ to be centered at the rest position $\mathtt{P}_0$ placed during the 3D warp initialisation stage (see Sec. 4.2).

5. Augmenting the Warp with New Points

The multiview 3D warps bring a novel procedure for image augmentation. Not only 3D warps allow for planar retexturing and augmentation as with standard 2D warps but they also permit to augment an image sequence by placing directly a new surface in the 3D metric space. Moreover, notice that by estimating the control points in 3D we detach the warping function from the camera projection. This implies that we can also arbitrarily modify the camera viewpoint and reproject the deforming surface. This results in a novel view synthesis of the deforming surface. In the next section we will present both algorithms for augmenting the warps in 2D and the novel concept of 3D augmentation.

5.1. Augmenting the warp in 2D

In general, if the control points are representative of the whole deformations, we may augment the estimated deforming shape with new samples. If we choose the first image frame, we consider a set of new points $\tilde{\mathtt{Q}}_1$ of size $\tilde{p} \times 2$. This new set of points are lying on the imaged surface. Thus the new 2D image measurement are given by:

$$\tilde{\mathtt{Q}} = \left[\; [\mathtt{Q}_1 \;\; \tilde{\mathtt{Q}}_1] \;\middle|\; \ldots \;\middle|\; [\mathtt{Q}_f \;\; \tilde{\mathtt{Q}}_f] \;\right]$$

of size $(p+\tilde{p}) \times 2f$. Given the known *warp transfer matrix* $\mathtt{L}$, $\mathtt{E}_\lambda$, and $\mathtt{P}$, now the problem is to find the estimate of the new point trajectories which gives the missing entries of $\mathtt{L}$ which we call $\tilde{\mathtt{L}}$. Thus by including $\tilde{\mathtt{L}}$ we have that:

$$\tilde{\mathtt{Q}} = \begin{bmatrix} \mathtt{L} \\ \tilde{\mathtt{L}} \end{bmatrix} \mathtt{E}_\lambda \left[\; \mathtt{P}_1 \;\middle|\; \ldots \;\middle|\; \mathtt{P}_f \;\right] \begin{bmatrix} \mathtt{M}_1 & \ldots & \mathtt{0} \\ \vdots & \ddots & \vdots \\ \mathtt{0} & \ldots & \mathtt{M}_f \end{bmatrix} \tag{11}$$

The problem is now to estimate the distance of the new 3D points from the control point centers. Such distances are stored in the unknown matrix $\tilde{\mathtt{L}}$ given by:

$$\tilde{\mathtt{L}} = \begin{bmatrix} \rho\left(d^2(\tilde{\mathbf{x}}_1, \mathbf{c}_1)\right) & \cdots & \rho\left(d^2(\tilde{\mathbf{x}}_{\tilde{p}}, \mathbf{c}_1)\right) \\ \vdots & \ddots & \vdots \\ \rho\left(d^2(\tilde{\mathbf{x}}_1, \mathbf{c}_l)\right) & \cdots & \rho\left(d^2(\tilde{\mathbf{x}}_{\tilde{p}}, \mathbf{c}_l)\right) \\ \tilde{\mathbf{x}_1} & \cdots & \tilde{\mathbf{x}}_{\tilde{p}} \end{bmatrix}, \tag{12}$$

where $\tilde{\mathbf{x}}_i$ for $i = 1 \ldots \tilde{p}$ are the unknown 3D entries given the added set of points. Instead of estimating the $\tilde{\mathbf{x}}_i$, we align the 3D points to the first frame giving x_i and y_i with $\tilde{\mathbf{x}}_i = (x_i\, y_i\, z_i)^\top$. We can write the cost function to optimise as a function of the unknown depth z_i:

$$\Gamma(z_1, \ldots, z_{\tilde{p}}) = \left\| \tilde{\mathtt{Q}}_1 - \tilde{\mathtt{L}}(z_1, \cdots, z_{\tilde{p}})\mathtt{E}_\lambda\mathtt{P}_1\mathtt{M}_1 \right\|_2^2$$

A single frame is not enough to obtain a reliable estimate of the points depth; thus we need at least more than 2 sets of matches for each new point. If we define the missing data mask $\mathtt{D}$ as in Eq. (8) we can solve for:

$$\Gamma(z_1, \ldots, z_{\tilde{p}}) = \left\| \mathtt{D} \odot (\tilde{\mathtt{Q}} - \tilde{\mathtt{L}}(z_1, \cdots, z_{\tilde{p}})\mathtt{E}_\lambda\mathtt{G}) \right\|_2^2$$

This optimisation can be performed with standard non-linear least squares. Initialisation for the depths $z_1, \ldots, z_p$ is obtained by averaging the depth of the b closest points to each new samples (b was fixed to 5.) Notice that this same optimisation of the *warp transfer matrix* $\tilde{\mathtt{L}}$ can be applied as well to the original $\mathtt{L}$ to obtain a global refinement of the warp matrices. Remember that the warp construction is based on an initialisation using rigid SfM. This further refinement may improve the whole estimate of the deformation in the case of unreliable initialisation of the warp.

5.2. Augmenting the warp in 3D

Given the known initial mean 3D shape $\mathtt{S}$ and control points displacements, we may directly augment the warp matrix $\mathtt{L}$ by inserting 3D points or even meshes belonging to complex 3D surfaces. Augmentation in 3D avoids to search for new feature correspondences in the image sequence. Also, the non-linear optimisation stage in Sec. 5.1 is not necessary since, by providing directly the set of points $[\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{\tilde{p}}]$, we can directly build the new $\tilde{\mathtt{L}}$ as in Eq. (12) and compute the new image point location $\tilde{\mathtt{Q}}$. Most interestingly is the fitting of a surface mesh using the mean points computed from the rigid SfM stage. Giving the set of points $[\mathbf{x}_1, \dots, \mathbf{x}_p]$ we can fit a 3D surface and reproject it back into the image plane as shown in the experimental section (Fig. 6). This augmentation is available regardless of the visibility of the surface at a given frame. In the 2D case, in order to augment the warp, it is compulsory to have a visible part of the projected shape where to match image points. Differently, by inserting the surface in 3D we may augment invisible parts (or highly occluded) and model their deformation using the learned 3D warp functions.

6. Experiments

6.1. Synthetic data

Our synthetic evaluation has the purpose to verify two different aspects of the proposed multiview 3D warp. First its deformation representative power is compared to classical NRSfM algorithms that are standard 3D reconstruction algorithms for sparse point sets. Then we show the algorithm resilience to increasing amount of missing data in the measurements. Notice that we can only do a sparse point evaluation since not all the standard methods for 3D modelling from images model directly dense deformation fields.

In order to understand the representative power of the warp we present two synthetic sequences with complex deformations. The *flag sequence* [10] is a motion captured cloth moving freely and displaying erratic and unsteady motion. We compare the results of our algorithm with Torresani et al.'s algorithm [18] (EM-PPCA), Bundle Adjustment (BA) [7], the BALM method [8] and two piecewise modelling approaches: Taylor et al.'s approach [17] using locally rigid motion (L-RIGID) and the piecewise quadratic method of Fayad et al. [10] (P-QUAD). The standard NRSfM approaches EM-PPCA, BA, and BALM achieved a 2D error of 19.02%, 16.01%, and 6.65% respectively[3]. Our approach does far better, 3.03%, almost reaching the best results of 1.50% and 2.04% provided by the piecewise SfM P-QUAD and L-RIGID respectively. The *MOCAP cylinder* instead presents a cylindrical shape showing strongly non-linear deformations. Our approach succeeds to obtain a 0.76% 2D error against the 0.55% error given by P-QUAD. In this test, given the sparseness of the points and the strong bending motion of the cylinder, it was not possible to apply successfully L-RIGID.

[3]The 2D error was defined as the Frobenius norm of the difference between the recovered 2D trajectories $\mathtt{Q}$ and the ground truth $\mathtt{Q}_{GT}$.

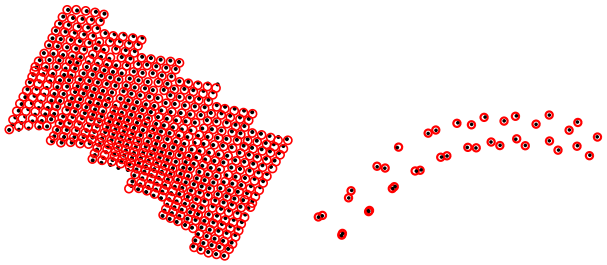


Figure 3. A selected frame from the flag and MOCAP cylinder sequences [10]. Black dots represent the 2D point ground truth while red circles shows the estimated points with the 3D warps.

A further test aims to compare our multiview 3D warp against increasing missing data with ground truth. The 3D warps are compared against standard algorithms for NRSfM since the available piecewise implementations P-QUAD and L-RIGID do not allow for missing data in the trajectories. For this test we used a 3D motion capture sequence of a real deforming face captured using a VICON system tracking and 37 points in 3D were then projected synthetically onto an image sequence of 74 frames using an orthographic camera model. Overall, we ran 100 tests for each configuration of missing data up to a 70% ratio.

In Fig. 4 we compare the results of our algorithm with Torresani *et al.*'s algorithm [18] (EM-PPCA), the BALM method [8] and Bundle Adjustment (BA) [7]. Our multiview 3D warps' behaved fairly well; the 2D error curve lies inbetween BA and EMPPCA/MP. The 2D error never goes over 5% and it shows a remarkable regularity at increasing levels of missing data.

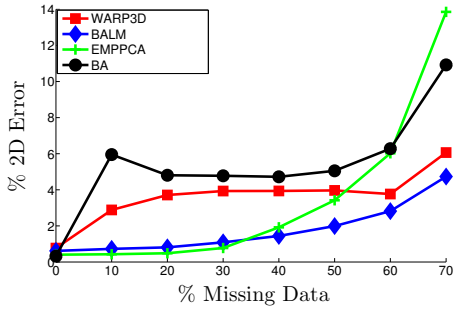


Figure 4. Synthetic test results showing a comparison between our multiview 3D warps and NRSfM methods.

6.2. Real data

We further test the 3D warps using two real image sequences. The first test was previewed in Fig. 1 and it

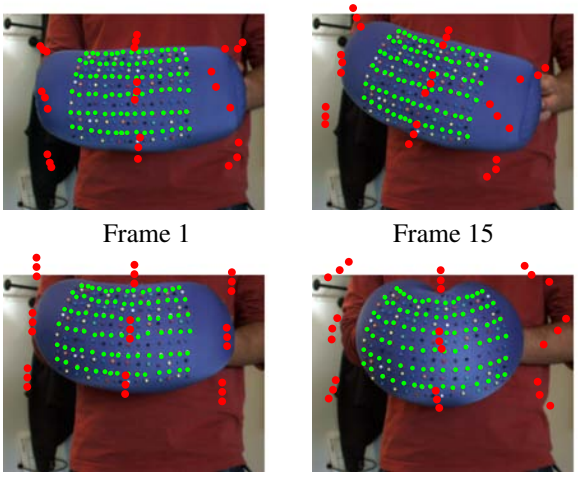


Figure 5. The four figures show the 27 control point displacement with red dots. The green dots shows the estimated 2D coordinates given the warp function.

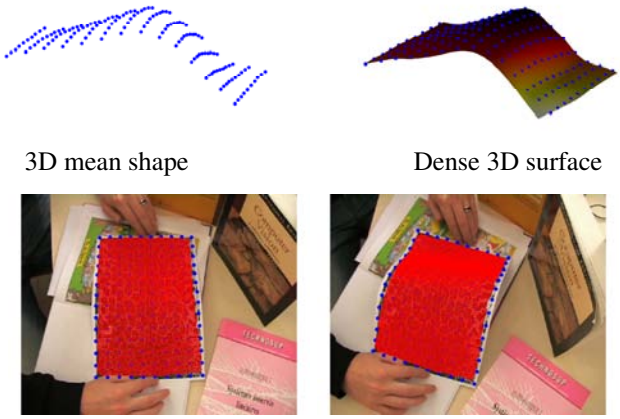


Figure 6. In the top left figure, we first show the sparse point stored in $\mathtt{S}$ obtained by rigid SfM. The top right figure shows an interpolated surface given the reconstructed 3D points. The bottom rows show two frames of the paper sequence. Notice here how the 3D mesh (in red) is correctly projected and bent into the image plane.

presents a moving (rotation and translation) cushion with bending deformations. A set of 90 image tracks were extracted and missing data were created synthetically with random sampling up to a 40% ratio. After the warp initialisation with 27 control points, we determine the variation in time of the control points and their relative camera projection matrices as presented in Fig. 5.

These frames were chosen with a purpose. Notice that in Frame 1 the grid is not perfectly regular as in Fig. 2 which shows the grid's initial configuration. This is because the mean shape which constructs the warp is computed from the whole sequence. Thus, already in the beginning of the sequence, the warp has to adapt to the configuration of the observed points. Frame 15 shows that when the shape moves almost rigidly the warp is projected and translated correctly by the estimated motion matrix $\mathtt{M}$. Frame 29 shows a configuration of the control points which is almost perfectly regular. In this frame, the mean 3D shape is very similar to the observed image data thus the algorithm does not need to deform much the control points. Frame 44 shows the strongest deformation in the sequence, notice how the control points are stretched to mimic the deformation. The average RMS 2D error given the known coordinates is 0.08 pixels.

6.3. Augmentation, cloning and retexturing

The capabilities of the *multiview 3D warp* is not only restricted to modelling performance. In the following we show three tasks that can be readily available only by using the learned 3D warp using the proposed framework.

Shape augmentation. We tested a further real sequence to present shape and image augmentation. The sequence is 105 frames long and it shows a paper bending and rotating. The warp was learned using 280 image points from which we reconstructed a 3D mean shape (top left image of Fig. 6.) Notably, we observed that the sparse 3D shape being bent and not at a resting position (i.e. flat) is not affecting negatively the whole algorithm performance. Given the sparse pattern, we augment the shape by surface interpolation obtaining a dense 3D mesh. The figures in the last row shows the reprojection of the deforming mesh into the image plane. Notice how the surface bending accurately describes the real image motion.

Deformation cloning. The learned deformation field is independent of the imaging conditions (i.e. the camera pose) and the shape motion. Thus the learned dense warp in the metric space can be easily reused to augment a new sequence or to transform the existing one. This was obtained by choosing an arbitrary 3×3 matrix $\hat{\mathtt{T}}$ such that $\mathtt{L}\mathtt{E}_\lambda \mathtt{P}_i \hat{\mathtt{T}} \mathtt{M}_i - \hat{\mathtt{Q}}_i$. The transformation $\hat{\mathtt{T}}$ is at the user discretion. In Fig. 7 we have chosen to scale the deformation by 2 and rotate it by $\alpha = 15$, $\beta = 15$, and $\gamma = 45$ degrees.

Image retexturing. Fig. 8 shows the retexturing of the paper image sequence where a synthetic texture is added to the bending paper. Notice that the augmentation is made by first projecting the dense mesh in Fig. 6 and then by retargeting the texture to augment the video with a logo.

7. Conclusion

We have proposed the multiview 3D warps. This new model improves state of the art in several ways. It is the first model to explicitly parameterize images of a deforming body by an average dense 3D shape and a set of 3D deformations combined with projection to 2D. Existing warps

Figure 7. Cloning the learned warp from the previous sequence (left image.) On the right we show the original surface in blue projected into the image frame. The smaller red mesh is cloned from the original but resized by two and rotated.

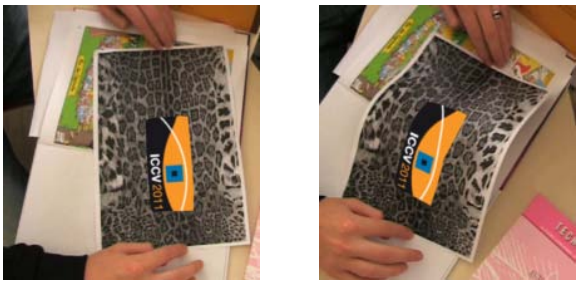


Figure 8. Automatic retexturing of the paper sequence.

map points from an image directly into another, and do not allow one to easily access the body's 3D shape and deformations, and camera pose, while NRSfM uses only sparse point matches and thus does not capture the deformation field for the whole 3D body. Our warps are templateless: they do not use specific priors on the observed body's shape. We have proposed a feature-based method that, from point matches between multiple images, estimates our multiview 3D warps' parameters. Thanks to our warps, points can be transferred from one image to another, and the captured body's deformations can be reused to augment the original images, to retarget the deformations or to alter the viewpoint. The reported experimental results on simulated and real data showed how image augmentation and retexturing could be performed. These results show that our warps' image modelling power is comparable or better to classical NRSfM methods. More importantly, we have experimentally demonstrated deformation cloning which was yet not possible using state of the art methods. We believe that our multiview 3D warps may open new research directions. In particular, the placement of the 3D control points and the estimation of the weights given to the various 3D shape priors are topics for which further research is especially important.

Acknowledgments. The authors would like to thank J. Fayad and L. Agapito for providing part of the data used in the synthetic and real experiments. We are also grateful to E. Muñoz for providing code for the retexturing test.

References

[1] A. Bartoli, M. Perriollat, and S. Chambon. Generalized thin-plate spline warps. *IJCV*, 88(1):85–110, May 2010. 1, 2, 3

[2] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI*, 11(6):567–585, 1989. 1, 2, 3

[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000. 1, 2, 4

[4] F. Brunet, A. Bartoli, R. Malgouyres, and N. Navab. NURBS warps. In *BMVC*, 2009. 2

[5] A. M. Buchanan and A. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *CVPR*, 2005. 4

[6] T. Cootes, C. Twining, V. Petrovic, K. Babalola, and C. Taylor. Computing accurate correspondences across groups of images. *PAMI*, 32(11):1994–2005, 2010. 1, 2

[7] A. Del Bue, F. Smeraldi, and L. Agapito. Non-rigid structure from motion using ranklet–based tracking and non-linear optimization. *IVC*, 25(3):297–310, March 2007. 2, 6

[8] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini. Bilinear factorization via augmented lagrange multipliers. In *ECCV*, 2010. 2, 5, 6

[9] J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *RAIRO Analyse Numérique*, 10:5–12, 1976. 2

[10] J. Fayad, L. Agapito, and A. Del Bue. Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In *ECCV*, 2010. 2, 6

[11] V. Gay-Bellile, A. Bartoli, and P. Sayd. Feature-driven direct non-rigid image registration. In *BMVC*, 2007. 1

[12] M. Marques and J. Costeira. Estimating 3D shape from degenerate sequences with missing data. *Computer Vision and Image Understanding*, 113(2):261–272, February 2009. 4

[13] S. Olsen and A. Bartoli. Implicit non-rigid structure-from-motion with priors. *Journal of Mathematical Imaging and Vision*, 31(2):233–244, 2008. 2

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

[15] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registation using Free-Form Deformations: Application to breast MR images. IEEE *Trans. on Medical Imaging*, 18(8):712–721, August 1999. 2

[16] M. Salzmann, J. Pilet, S. Ilic, and P. Fua. Surface deformation models for nonrigid 3D shape recovery. *PAMI*, 29(8):1–7, August 2007. 1, 2

[17] J. Taylor, A. Jepson, and K. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, pages 2761–2768, 2010. 2, 6

[18] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *PAMI*, pages 878–892, 2008. 2, 6

[19] T. Vercauteren, X. Pennec, A. Perchanta, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1), March 2009. 1, 2