Balloon shapes: reconstructing and deforming objects with volume from images

Sara Vicente Lourdes Agapito
School of EECS, Queen Mary, University of London
[sara.vicente, lourdes.agapito]@eecs.qmul.ac.uk

***Abstract*—Reconstructing the shape of a deformable object from a single image is a challenging problem, even when a 3D template shape is available. Many different methods have been proposed for this problem, however what they have in common is that they are only able to reconstruct the part of the surface which is visible in a reference image.**

In contrast, we are interested in recovering the full shape of a deformable 3D object. We introduce a new method designed to reconstruct closed surfaces. This type of surface is better suited for representing objects with volume.

Our method relies on recent advances in silhouette based reconstruction methods to obtain the template from a reference image. This template is then deformed in order to fit the measurements of a new input image. We combine an inextensibility prior on the deformation with powerful image measurements, in the form of silhouette and area constraints, to make our method less reliant on point correspondences.

We show reconstruction results for different object classes, such as animals or hands, that have not been previously attempted with existing template methods.

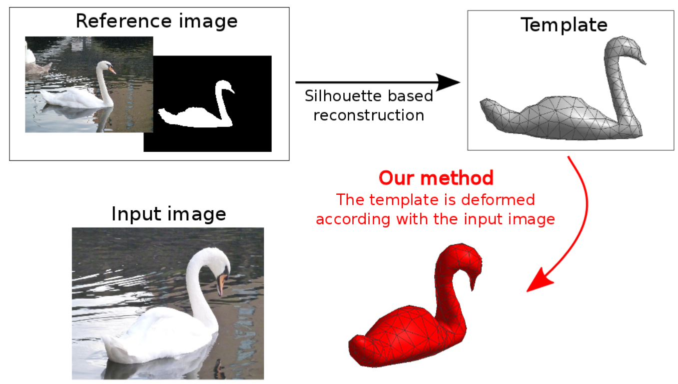

Figure 1. Overview of our method for template based reconstruction of deformable objects. We start by obtaining a template from a reference image using silhouette based methods [16]. Given a new input image the template shape is deformed in order to fit the image measurements. Our method recovers the shape of a deforming object based on only two images and is able to reconstruct the parts of the surface which are not visible in the reference image.

I. Introduction

Reconstructing the 3D shape of an object from images is a longstanding challenge in computer vision. While for rigid objects there are stable methods that have been used in practical applications, such as the preservation of cultural heritage by constructing 3D models of important artefacts, for deformable objects this is not the case.

Traditional non-rigid structure from motion methods (NRS*f*M) are designed to reconstruct deformable objects from 2D correspondences in a video [5]. They represent the object as a linear combination of basis shapes. This basis is recovered together with the deformation coefficients and rigid motion. However, NRS*f*M methods have a few important restrictions in practice: (1) in the type of objects they can reconstruct, since they are better suited to reconstruct small deformations of a rigid object, such as a talking face; (2) in requiring clean point correspondences and not being resilient to missing data; and (3) in requiring a considerable amount of out of plane rotation.

One way to overcome the limitations of NRS*f*M is to manipulate the acquisition system and increase the amount of information available, by using, for example, multiple synchronized views of the object [25] or 3D scanner data [28].

This research was funded by the European Research Council under the ERC Starting Grant agreement 204871-HUMANIS

Using information about the shape of the object is another successful alternative previously considered. This is the direction we take in this paper. In particular, we focus on reconstruction from a single image of a deformable object, given a 3D template of its undeformed shape. Previous methods that fit a deformable shape to a single image can be roughly divided into two groups, depending on the amount of information available about the shape of the object: model based methods and template based methods.

Model based methods provide a low dimensional parametrization of both the shape and deformations of an object class. Models of this type are available for classes such as faces [3] or the human body [1]. When used for reconstruction from a single image, these methods jointly infer the shape of the specific instance of the object depicted in the image and the pose or deformations observed. Very high quality reconstructions from single images have been obtained using such methods (see Fig. 2 (a)). However, they require large amounts of 3D scans to learn the parametrized model, making it prohibitive to build such a model for all possible object classes.

Template based methods assume that a 3D template shape corresponding to a reference image is known. They provide a simpler alternative to model based methods, since they do not require a training stage, only this template

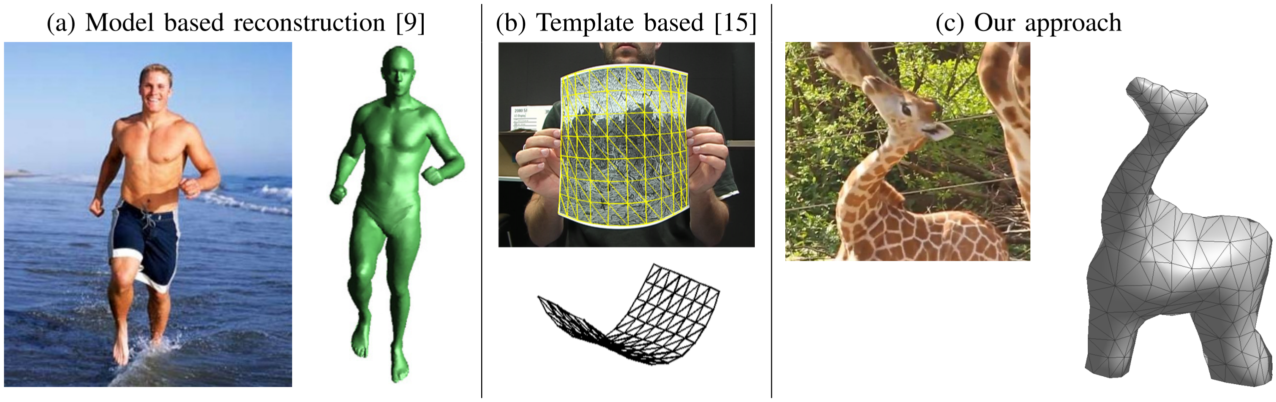

Figure 2. Comparison of single image reconstruction results obtained with model based methods (a), template based methods (b) and our method (c). While model based methods have been successfully used to reconstruct the complete shape of an object, such as the human body (a), existing template based methods focus on reconstructing surfaces with boundaries (b) that are not suitable to represent objects with volume (c). In contrast, our template based method can cope with reconstruction of closed surfaces, narrowing the gap between the type of objects considered by template and model based methods. The template used for our reconstruction can be seen in Fig. 5. The result of model based methods was reproduced from [9] and the result of the template based method was reproduced from [15].

shape. The template being associated with a reference image gives a further advantage: obtaining point correspondences between the template and the input image can be done using established point correspondence methods for 2D images. This is not the case for model based methods. Although, the template provides no information about the space of deformations, geometric constraints are used to restrict this space. The most commonly used geometric constraint is a near-isometry (or inextensibility) constraint, i.e. the deformations should preserve geodesic distances between points in the surface. This has been shown to be a sensible constraint for many shapes [6], including the human body.

Despite these advantages, existing template based methods have only been demonstrated in much more constrained scenarios with less impressive results, when compared with model based methods (see Fig. 2). The goal of this paper is to narrow this gap by showing that template based methods can be used to reconstruct more challenging objects than previously considered.

Our main observation is that existing methods are restricted to templates which consist of a surface (or mesh) with a boundary, and that this surface is fully visible in the reference image. Instead, we consider templates which are closed surfaces without boundaries (we refer to them simply as *closed surfaces* hereafter). This type of surface encloses a 3D volume and is therefore more suitable to represent a larger class of deformable shapes, such as animals, a hand or the human body. Closed surfaces pose a problem to existing methods since they inevitably generate self-occlusions. Although, some methods can cope with occlusions [18], they are only present in the input image, while the reference image necessarily depicts the template fully unoccluded.

Our method is divided into two steps and Fig. 1 gives an overview. The first step is the template acquisition from the reference image. Previous methods use as template shape: planar shapes [21], [15], easily obtained from a fronto-parallel reference image, rigid reconstruction from a few frames [2], or a shape obtained with an RGB-D camera [15]. Instead, we rely on our extension of recent advances in silhouette based single view reconstruction [16] to obtain a 3D template from the reference image.

In the second step, the closed surface template is deformed to fit the input image, by penalizing deformations which are not isometric. We make use of three different types of image cues: silhouette and area constraints and point correspondences. Silhouette constraints are a powerful source of information that have been used for different applications: multi-view stereo [25], model based single image reconstruction [24], [9] or building morphable models from 2D images [7]. The use of the silhouette and area constraints allows to reduce considerably the number of point correspondences needed.

In practice, our method is able to reconstruct a deforming object using the information available in only two images: the reference and the input image.

II. Related work

There is an extensive list of template based methods for single view reconstruction and we refer the interested reader to [21] for a recent review. Typically, template based methods rely on either sparse [21], [15], [18] or dense correspondences [2] between the reference and the input images, and use inextensibility constraints as the geometric prior to define the space of deformations [21], [15], [18], [2].

Existing methods differ mostly in the way the problem is formulated and in the optimization method used. Different convex formulations [21] as well as a closed form solution [2] have been proposed. However, they have in common the application scenarios and the type of surfaces that can be reconstructed: surfaces that are fully visible in the reference image, particularly developable surfaces, i.e. surfaces that are isometric to a plane.

Model based methods differ from template based methods in that they provide information about an object *class*, as opposed to a specific instance of a class. Therefore, the representation of the model accounts for both changes in shape and changes in pose. For example, the SCAPE model [1], a human body model that has been used for reconstruction from single images [9], models the variations in shape as a linear combination of basis shapes and the pose as an articulated model. The biggest disadvantage of these methods is the need for a considerable amount of 3D data in order to learn the model, e.g. 107 3D scans of humans were used for the SCAPE model in [1] and 200 scans of different people to build the face model in [3].

In order to overcome this difficulty, the method in [7] shows how to build a morphable model from considerably less data which can be easily obtained. To build a model for dolphins, they only required: a single 3D template, 32 images and six point correspondences per image. Similarly to our approach, they make use of silhouette information as well as a few point correspondences. However, this method is not designed to reconstruct from a single image, since the representation of the deformations needs to be learned from multiple images.

Our method can alternatively be seen as a mesh deformation method, where the deformations are guided by image measurements. Mesh deformation methods are extensively studied in the graphics community (e.g. see [4] for a review on existing linear deformation methods). When deforming a mesh, their goal is to preserve some of the intrinsic properties of the shape, such as its local structure. They differ from our method on what “guides” the deformation, since they typically rely on precise user interaction in 3D space. Related to our approach is [12], a mesh deformation method that is guided by 2D contours. These contours are carefully traced by hand and can represent either the silhouette of the object or internal edges. Obtaining these high quality internal contours automatically from images is not trivial.

The approach most closely related to ours [8] is a template based reconstruction method applicable to closed surfaces that penalizes deformations that do not preserve geodesic distances between points. This method can be used for 3D shape denoising and reconstruction from 2D projections. However, it requires multiple views of the deformed object, no occlusions and a large amount of point correspondences.

Our method is able to reconstruct a deformed shape from a single input image. The use of silhouette constraints makes it less reliant on a large amount of point correspondences. Furthermore, since we use silhouette based methods to obtain the template from the reference image, effectively our method recovers an object in two different poses from only two views. We use it to reconstruct the full 3D shape of interesting object classes, such as animals, that have not been previously considered by other template based methods.

III. Obtaining a template from the reference image

We rely on a silhouette based reconstruction method to obtain our template shape from the reference image[1]. Such methods have a long tradition in computer vision: a symmetry seeking method was introduced in 1988 for reconstruction of tube like shapes from a single image [26]. Furthermore, sketch based modeling techniques based on inflation, such as *Teddy* [11] and *Fibermesh* [14], are very popular in the graphics community.

In this paper, we use a modification of the silhouette based method in [16]. Similarly to other inflation techniques [11], [14], it recovers a balloon-like surface that respects the silhouette of the object. It requires that the silhouette of the object is traced in the image, i.e. it requires the object to be segmented. We use grabcut [19] to segment the reference image.

The reconstructions obtained with this method are only plausible if the object and its depiction in the image have certain properties: the object should have a plane of symmetry parallel with the image plane and it should not have concavities that cannot be inferred from its silhouette. For example, if the image depicts a side view of a standing dog, the reconstruction would not recover correctly the legs: the two front legs would be joined together in a single leg and the same would happen for the two back legs, since the fact that they are separated is not visible from a side view silhouette. Although, these constraints limit the type of objects that can be reconstructed and also the choice of reference image, in the experimental section we show that this method is still applicable to a wide variety of objects. Note that, in the experiments we also show results for templates that are not obtained using this method, since our method is independent of the way the template is recovered.

Nevertheless, we advocate silhouette based reconstruction methods because they provide a simple automated way of obtaining a closed surface template from the reference image. For completeness, we describe in more detail the method in [16] and our extension to it.

Let $\Omega \subset \mathbb{R}^2$ be the image plane and $S \subset \Omega$ the object segmentation, i.e. S contains the points inside the silhouette of the object. The goal is to recover a height function $u : S \rightarrow \mathbb{R}$, assigning a depth value $u(x, y)$ to all points $(x, y) \in S$.

In [16], the height function u is obtained as the minimum

[1]This type of method is also commonly named *single view reconstruction*. We refer to it as *silhouette based reconstruction*, to emphasize the difference with our template based single image reconstruction method.

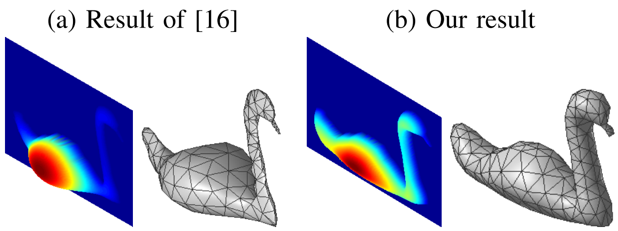

Figure 3. Comparison of the silhouette based reconstructions obtained with [16] (a) and with our method (b). The result is shown both as a height function and a coarse mesh. The reconstruction of [16] does not recover properly some of the thin structures of the object, while using our extended energy correctly inflates the neck. These reconstructions correspond to the image in Fig. 1.

of the energy function:

$$E(u) = \int_S \sqrt{1 + |\nabla u|^2} \quad dxdy \tag{1}$$

subject to a volume constraint of the form $\int_S u \ dxdy = V$, for a constant V. The intuition for this energy is that the height function u should have minimal area for a fixed enclosing volume V, which is an image dependent parameter. This is consistent with the procedure of inflating a balloon while forcing it to respect a silhouette.

A disadvantage of this method is that it does not inflate properly thin structures on the object, for instance the neck of the swan in Fig. 3 (a). In order to encourage that thin structures have some volume, we add an extra term to the energy. Our new energy has the form:

$$E(u) = \int_S (u - f)^2 + \sqrt{1 + |\nabla u|^2} \quad dxdy \tag{2}$$

subject to the same volume constraint: $\int_S u \ dxdy = V$.

The first term acts as a shape prior similar to the one used in [17]:

$$f(x, y) = min\{\mu_{\text{cutoff}}, \mu_{\text{offset}} + dist((x, y), \partial S)\} \tag{3}$$

where μ_{cutoff} and μ_{offset} are user selected weights and $dist((x, y), \partial S)$ corresponds to the distance of the interior points to the closest point in the silhouette. Function f can be seen as a guess on the shape of the object and it encodes the generic assumption that objects get thicker the further away the point is from the silhouette, while encouraging that thin structures are still inflated.

Fig. 3 (b) shows the effect of adding this term to the energy. We use the optimization method in [16] to minimize the energy function (2), an iterative gradient descent method with a projection step to impose the volume constraints. Since the function remains convex with the addition of the extra term, we obtain the global optimal solution.

Given the optimal height function u, we combine it with a reflected version of it to create a closed surface. We then use the algorithm and reference implementation of [27] to create a coarser and more uniformly spaced mesh. Fig. 3 (b) shows an example of the final template mesh created using this procedure.

IV. Template based reconstruction of a deformable object from a single image

After obtaining a template shape using the silhouette based reconstruction method, the goal is to reconstruct a deformed shape consistent with a new image.

The template mesh consists of: the set of vertices $\mathcal{V}$ with corresponding 3D coordinates, $\mathcal{T} = \{\bar{\boldsymbol{x}}_p, p \in \mathcal{V}\}$ with $\bar{\boldsymbol{x}}_p = (\bar{x}_p, \bar{y}_p, \bar{z}_p)$; the set of edges $\mathcal{E}$; and the set of triangular facets.

The goal is to infer the coordinates of the deformed mesh, $\mathcal{X} = \{\boldsymbol{x}_p, p \in \mathcal{V}\}$ with $\boldsymbol{x}_p = (x_p, y_p, z_p)$, by minimizing an energy function of the form:

$$E(\mathcal{X}) = \lambda_p E^{prior}(\mathcal{X}) + E^{data}(\mathcal{X}) \tag{4}$$

where λ_p is a user defined weight. This energy encourages the shape to deviate little from the template and to fit the image measurements.

A. Shape prior

The first term of the energy (4) encodes an inextensibility constraint, similar to previous template based methods [21], [18], penalizing deformations that do not preserve the euclidean distance between nearby points. This is a commonly used approximation to the more principled constraint of preserving geodesic distances between all possible pairs of points.

We define a set of edges, $\bar{\mathcal{E}}$, which includes the mesh edges, $\mathcal{E}$, as well as additional edges connecting vertices $p, q \in \mathcal{V}$ which have a mesh neighbour in common:

$$\bar{\mathcal{E}} = \mathcal{E} \cup \{(p, q) \mid \exists r \in \mathcal{V} : (p, r), (q, r) \in \mathcal{E}\} \tag{5}$$

The shape prior is defined as:

$$E^{prior}(\mathcal{X}) = \sum_{(p,q) \in \bar{\mathcal{E}}} (dist(\boldsymbol{x}_p, \boldsymbol{x}_q) - dist(\bar{\boldsymbol{x}}_p, \bar{\boldsymbol{x}}_q))^2 \tag{6}$$

where $dist(\boldsymbol{x}, \boldsymbol{y})$ corresponds to the 3D euclidean distance between the points $\boldsymbol{x}$ and $\boldsymbol{y}$.

Preventing volume changes The shape prior in (6) does not prevent unnatural volume changes between the template mesh and the deformed mesh. This is undesirable for the type of objects we consider, since the deformations should not have an impact on the volume of the mesh.

Similarly to [29], in order to prevent those changes, we add extra vertices inside the mesh. The form of the energy remains the same, but the set of variables is extended to include those extra vertices. The set of edges is also extended to include both intra-connections between these interior points and connections with mesh vertices. The addition of these interior vertices should help to maintain the overall structure of the shape.

(a) Internal structure (b) Chamfer distance $D(x, y)$

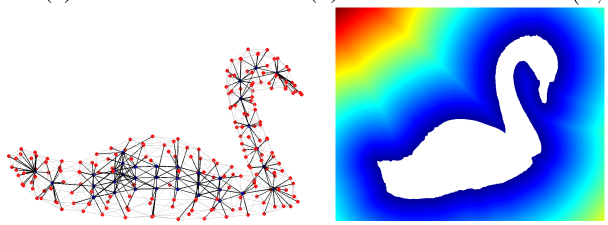

Figure 4. (a) An example of the internal structure added to the mesh to help preserve the volume. The nodes in blue are the new set of internal nodes, the nodes in red are the mesh nodes, edges in black are the new edges connecting two internal nodes or an internal node to a mesh node, and grey edges are the mesh edges. (b) Illustration of the Chamfer distance $D(x, y)$. Red corresponds to higher values of $D(x, y)$, blue to small values and white to zero.

Fig. 4 (a) shows an example of the vertices and edges added to the mesh in Fig. 3 (b). We follow a fully automatic heuristic procedure similar to [29] to obtain these new vertices and edges. The vertices considered are either in the skeleton of the shape or in a regular grid fitted inside the mesh. They are chosen such that their distance is similar to the average length of the mesh edges. The extra edges connect nearby internal vertices and each mesh vertex with its closest internal vertex.

Furthermore, when using a template obtained with the method in Section III, we add additional edges connecting vertices which are located at symmetric positions with respect to the image plane.

Coping with changes in scale For some of our experiments, the scale of the object changes from the template to the input image. To cope with changes in object scale, we add an additional variable to the energy and modify the prior term:

$$E^{prior}(\mathcal{X}, sc) = \sum_{(p,q)\in\bar{\mathcal{E}}} (dist(\boldsymbol{x}_p, \boldsymbol{x}_q) - sc \times dist(\bar{\boldsymbol{x}}_p, \bar{\boldsymbol{x}}_q))^2$$

The new variable sc is estimated together with the shape coordinates $\mathcal{X}$.

B. Data term

The second term of the energy (4) encourages reconstructions that are in agreement with the image measurements, i.e. reconstructions that reproject correctly into the image. We assume an orthographic camera and align the world coordinates with the camera referential. The data term is of the form:

$$E^{data}(\mathcal{X}) = \lambda_s E^{silh}(\mathcal{X}) + \lambda_a E^{area}(\mathcal{X}) + \lambda_r E^{repr}(\mathcal{X}). \tag{7}$$

Silhouette and Area constraints Given the input image that we want to reconstruct, we start by segmenting the object using grabcut [19].

We use two types of data terms based on this segmentation: $E^{silh}(\mathcal{X})$, a silhouette term that encourages points to reproject inside the silhouette, and $E^{area}(\mathcal{X})$, an area term that penalizes differences between the area inside the silhouette and the area of the projected shape. Similar terms have been used before in model-based single view reconstruction [24], [9]. The first term has the form:

$$E^{silh}(\mathcal{X}) = \sum_{p\in\mathcal{V}} (D(x_p, y_p))^2 \tag{8}$$

where $D(x, y)$ is the Chamfer distance map, positive outside the silhouette and 0 inside the silhouette (see Fig.4 (b) for an illustration). This term penalizes points that reproject outside the silhouette.

The second term has the form:

$$E^{area}(\mathcal{X}) = (Area(\mathcal{X}) - A)^2 \tag{9}$$

where A is the target area, i.e. the area enclosed by the silhouette in the input image, and $Area(\mathcal{X})$ approximates the 2D area of the reprojection of the shape and it is defined as:

$$Area(\mathcal{X}) = \sum_{t\in\mathcal{S}} \left| x_1^t(y_2^t - y_3^t) + x_2^t(y_3^t - y_1^t) + x_3^t(y_1^t - y_2^t) \right| \tag{10}$$

where $\mathcal{S}$ is the set of visible mesh triangles, i.e the mesh triangles that have all the vertices visible, $(\boldsymbol{x}_1^t, \boldsymbol{x}_2^t, \boldsymbol{x}_3^t)$ are the three vertices associated with triangle t and $\boldsymbol{x}_i^t = (x_i^t, y_i^t, z_i^t)$, $i = 1, 2, 3$ are the coordinates of each of the vertices.

These two terms complement each other. The silhouette term alone can leave parts of the foreground segment that are not properly filled in, since the fact that no point reprojects on them is not penalized. However, such a solution would be penalized by the area term.

Note that our silhouette term does not require knowing the contour generators differing from [22], [29], [12], [7]. While [22], [29] assumes that the contour generators are known (either manually traced by a user [29], or correspond to the boundary of the mesh [22]), the methods in [12], [7] explicitly find the contour generators in an iterative fashion and use them to "pull" the shape towards the silhouette. Finding the contour generators is a non-trivial step and it is only feasible if the template shape is properly rotated and aligned with the input image [12], [7]. We bypass this step by using a different formulation of the silhouette constraint.

Point reprojection constraints Similarly to previous template based methods [21], [2], we also use point reprojection constraints, penalizing specific points in the shape that do not reproject onto a given 2D point. To enforce such constraints, we require point correspondences between the 3D shape and the input image. If the template shape is associated with a reference image, these point correspondences can be computed using descriptor based matching methods, such as SIFT [13]. Notice that, in contrast with most previous methods, we require very few point correspondences.

The point correspondences associate the 2D point $(\bar{u}_i, \bar{v}_i)$

in the reference image with the point (u_i, v_i) in the input image, with $i = 1, ..., P$ and P the total number of correspondences. For each of these correspondences, we compute the barycentric coordinates of the corresponding 3D point in the template mesh. This 3D point is written as a function of three mesh vertices in the following way: $a_i\bar{\boldsymbol{x}}_1^i + b_i\bar{\boldsymbol{x}}_2^i + c_i\bar{\boldsymbol{x}}_3^i$.

Finally, the energy term that encodes the point reprojection constraints has the form:

$$E^{repr}(\mathcal{X}) = \sum_{i=1}^{P}((a_i x_1^i + b_i x_2^i + c_i x_3^i) - u_i)^2 + ((a_i y_1^i + b_i y_2^i + c_i y_3^i) - v_i)^2 \quad (11)$$

Optimization We use a standard non-linear least squares solver, the Levenberg-Marquardt algorithm for optimizing energy (4). The template is used as initialization.

V. EXPERIMENTS

We use the method described in the previous sections for reconstruction of deformable objects from a single image. Since, to the best of our knowledge, there is no ground truth dataset that fits our requirements, i.e. with a closed mesh as template, we rely on a qualitative evaluation of our method. However, we do compare the results of our approach with [7] on their dolphin dataset.

Reconstruction from a template obtained with silhouette based methods For our first set of experiments we rely on the method described in Section III to obtain the template from the reference image. We show results of this experiment in Fig. 5. Our method is able to reconstruct challenging deforming shapes from a single image, such as a hand.

The template shapes can be seen in Fig. 5 (b) and they show the potential of the silhouette based reconstruction method for obtaining a template from the reference image for a wide variety of objects. One of the limitations of this method is visible for the giraffe, where the method fails to recover the four legs, grouping them instead into two.

In Fig. 5 (a) we show the locations of the point correspondences for each image. We use two types of point correspondences: obtained automatically with SIFT (yellow) and manually placed (magenta). The total number of correspondences used is 23, 193, 1 and 1 respectively. The number of correspondences required by our method is significantly lower than previous template based methods. Notice that for the giraffe image, despite the large amount of correspondences, they are concentrated on the torso, not on the deforming neck. Our method is able to recover that deformation using the information provided by the silhouette.

The running time for these examples was between 8 and 49 seconds per image, using Matlab's non-linear least squares solver.

The dolphin dataset was introduced in [7] and it was used to recover a full morphable model for this class. We use this dataset to show the flexibility of our method in using a template shape that is not obtained with the method in Section III. We use the same template as in [7] with the following modifications: the mesh provided by [7] has only 100 vertices and it is used in conjunction with an interpolation rule to create a smooth surface. We use Loop subdivision to obtain this smoother mesh and coarsen it using [27]. The final mesh has 300 points and is shown in Fig. 6 (a).

We have used both the silhouettes and the point correspondences provided by [7]. Fig. 6 shows two reconstructions obtained with our method and with [7]. Our method is unable to recover shape variations, that are instance specific. This can be seen on the first row: the tail is thinner in our reconstruction, since we keep its shape from the template. Nevertheless, our method recovers correctly the deformations in terms of pose and, in contrast with [7], can reconstruct the shape from a single image.

Initialization The non-linear least squares optimization method we use requires an initialization. We use the template shape as initialization after automatically centring it with respect to the input image, by translating it.

When there are significant changes in viewpoint from the template to the input image, e.g. for the last result in Fig. 5 and for the dolphins in Fig. 6, we use the optimization algorithm twice for a different choice of weights λ. In the first run, the weight λ_p is larger than the other weights, which favours rigid transformations of the template shape. This recovers a rough alignment of the template with the input image. In the second run, we initialize with the result of the first run and increase weights corresponding to the data term to obtain a better fit to the data. Similar strategies have been previously used for example in [9].

Scale We have included the scale variable described in Section IV-A in the optimization for the dolphin dataset, since there are considerable changes in scale between the template and the input images. The final scale for the examples in Fig. 6 is 2 and 2.4 respectively.

Limitations of the orthographic camera model Since the method assumes an orthographic camera there is an inherent single per-image flip ambiguity that cannot be resolved automatically and must be done manually, similarly to all other orthographic reconstruction methods.

In addition, since our method reconstructs 3D shape from a single image, it shares the ill-posedness common to other single view reconstruction problems. In particular, for articulated shapes it has the same "forward-backward" local ambiguity observed in 3D human pose estimation from a single view [23]. We observed this ambiguity only for the mannequin in the third row of Fig. 5: a solution where the leg is pushed forward and the arm backwards has similar

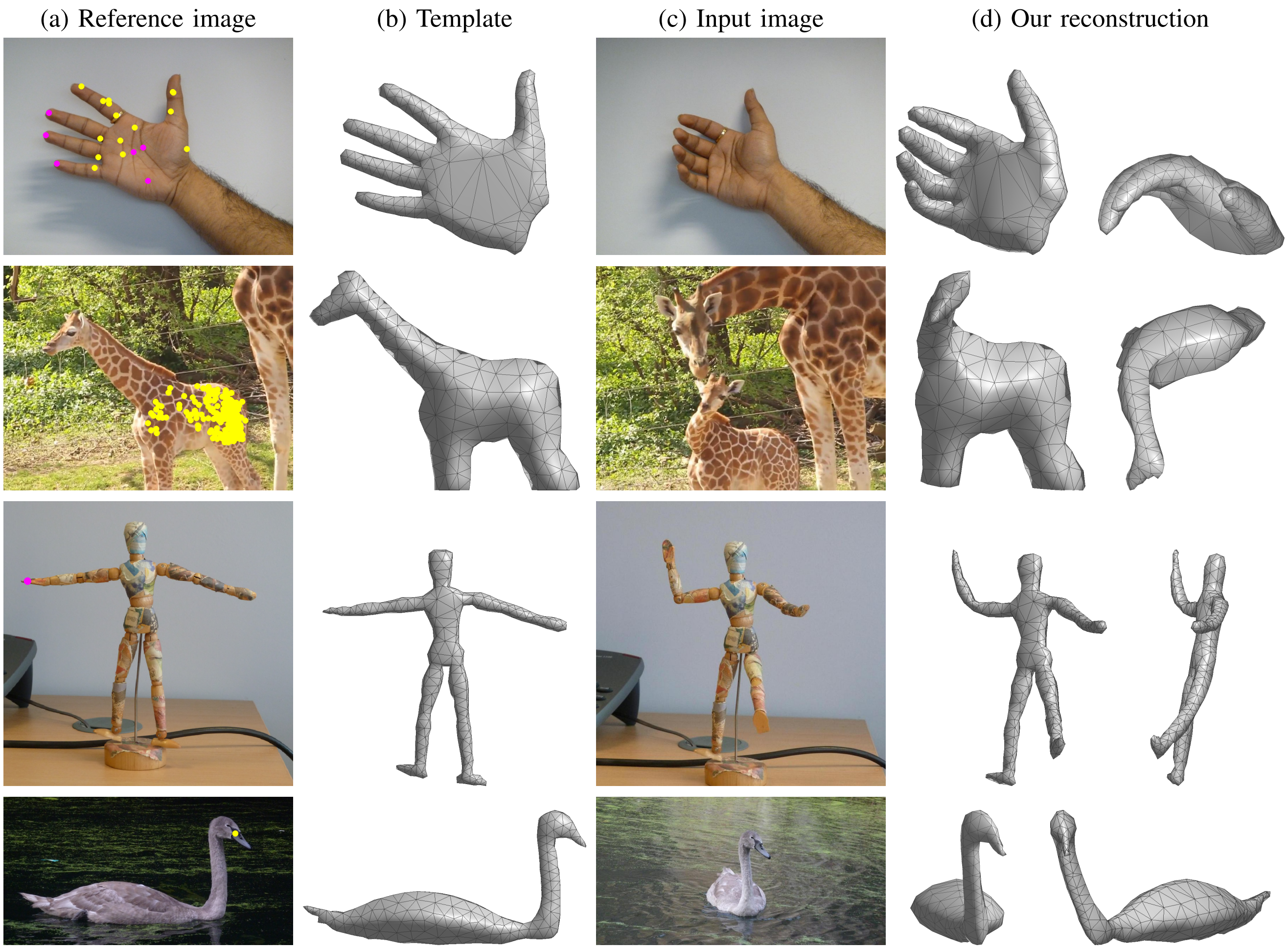

Figure 5. Results of our method when the template is obtained with silhouette based methods. We show two views of our reconstruction in (d): a view aligned with the camera and a top view (for the first two images) or a view corresponding to rotating the camera 45 degrees to the right (for the last two images). In (a) we also show the location of the point correspondences for each image and their origin, either automatically obtained with SIFT (yellow) or manually placed (magenta).

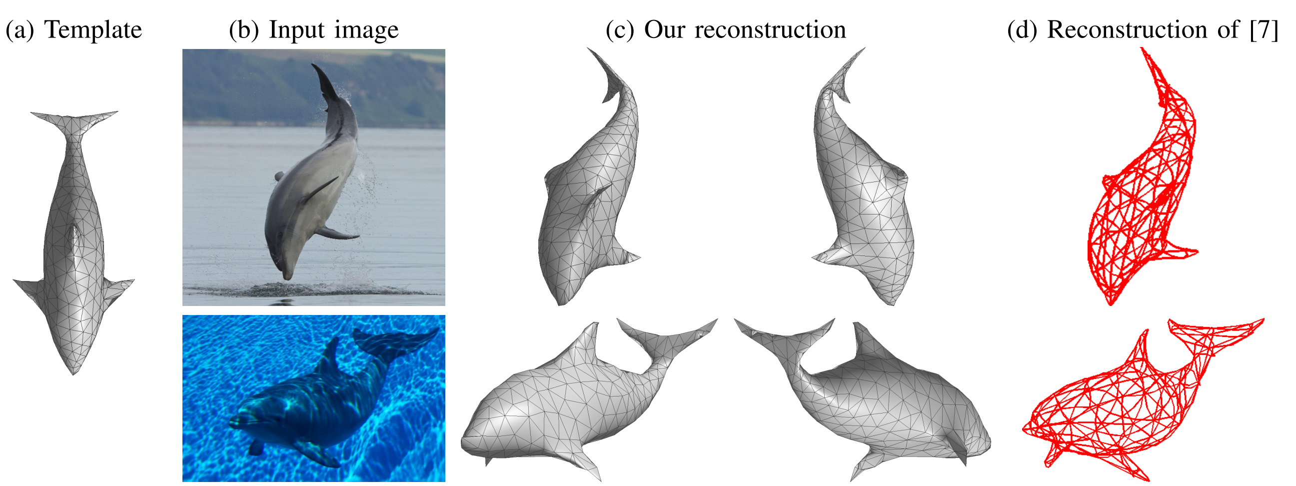

Figure 6. Results for the dolphin dataset. We use the images, silhouettes and point correspondences provided by [7]. We show two views of our reconstruction in (c): a view aligned with the camera and after rotating 180 degrees. For comparison, in (d) we show results produced with the reference implementation of [7] and the visualization tool therein. This experiment shows that our method can cope with templates from a source different than silhouette based reconstruction. The first image in (b) is copyright of Peter Asprey.

(but higher, for this particular case) energy.

This type of local flip ambiguity cannot be resolved without the use of extra-information, such as: (1) introducing extra views with some camera motion, which can be available in the case of video sequences; (2) allowing extra user interaction to select the correct local flip; or (3) using motion priors for some particular shapes.

Automation Our method requires some limited user interaction: in choosing the weights μ_i and λ_i and the volume V, in performing interactive segmentation of both images and, in certain cases, providing a small number of additional manual point correspondences.

We have observed that the silhouette based reconstruction and our method are quite resilient to changes in the weights, leaving room to some degree of automation. We could also rely on cosegmentation methods [20] to segment the same object in both images, or use a similar strategy to [9] alternating between estimating the shape and the segmentation.

These possible automations still leave open the problem of selecting a reference image. Recall that the silhouette based method described in Section III assumes that the reference image is aligned with a plane of symmetry of the object. Selecting a suitable reference image from a video is an interesting future direction of research. Another possible extension of our work is to update the template shape with information from multiple images, allowing to recover from the inaccuracies of the template reconstruction step.

VI. Conclusion

In this paper we have introduced a new method for template based reconstruction. Our method is designed to reconstruct closed surfaces, which are more suitable to represent objects with volume. It combines an inextensibility prior with a data term dependent on point correspondences, silhouette constraints and an area constraint. In the experiments we show that our method is able to reconstruct challenging shapes that have not been previously considered by other template based methods.

We believe that our method's ability to reconstruct the full 3D shape of a deformable object in two different poses using only two images complements recent efforts in using 3D shape information for object recognition [10], particularly for classes of deformable objects such as animals.

References

[1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM Trans. Graph.*, 2005.

[2] A. Bartoli, Y. Grard, F. Chadebecq, and T. Collins. On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces. In *CVPR*, 2012.

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

[4] M. Botsch and O. Sorkine. On linear variational surface deformation methods. *IEEE TVCG*, 2008.

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

[6] A. Bronstein, M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer, 2008.

[7] T. Cashman and A. Fitzgibbon. What shape are dolphins? building 3d morphable models from 2d images. *PAMI*, 2012.

[8] Y. Devir, G. Rosman, A. Bronstein, M. Bronstein, and R. Kimmel. On reconstruction of non-rigid shapes with intrinsic regularization. In *NORDIA*, 2009.

[9] P. Guan, A. Weiss, A. Balan, and M. Black. Estimating human shape and pose from a single image. In *CVPR*, 2009.

[10] D. Hoiem and S. Savarese. *Representations and techniques for 3D object recognition and scene interpretation*, volume 15. Morgan & Claypool Publishers, 2011.

[11] T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: a sketching interface for 3d freeform design. In *ACM SIGGRAPH*, 1999.

[12] V. Kraevoy, A. Sheffer, and M. van de Panne. Modeling from contour drawings. In *Eurographics Symposium on Sketch-Based Interfaces and Modeling*, 2009.

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004.

[14] A. Nealen, T. Igarashi, O. Sorkine, and M. Alexa. Fibermesh: designing freeform surfaces with 3d curves. In *ACM Trans. Graph.*, 2007.

[15] J. Östlund, A. Varol, and P. Fua. Laplacian meshes for monocular 3d shape recovery. In *ECCV*, 2012.

[16] M. Oswald, E. Toppe, and D. Cremers. Fast and globally optimal single view reconstruction of curved objects. In *CVPR*, 2012.

[17] M. R. Oswald, E. Töppe, K. Kolev, and D. Cremers. Non-parametric single view reconstruction of curved objects using convex optimization. In *DAGM*, 2009.

[18] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *IJCV*, 2011.

[19] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Trans. Graph.*, 2004.

[20] C. Rother, V. Kolmogorov, T. Minka, and A. Blake. Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs. In *CVPR*, 2006.

[21] M. Salzmann and P. Fua. Deformable surface 3d reconstruction from monocular images. *Synthesis Lectures on Computer Vision*, 2010.

[22] M. Salzmann and R. Urtasun. Beyond Feature Points: Structured Prediction for Monocular Non-rigid 3D Reconstruction. In *ECCV*, 2012.

[23] C. Sminchisescu. 3d human motion analysis in monocular video: techniques and challenges. *Human Motion*, 2008.

[24] C. Sminchisescu, A. Telea, et al. Human pose estimation from silhouettes. a consistent approach using distance level sets. In *WSCG International Conference for Computer Graphics, Visualization and Computer Vision*, 2002.

[25] J. Starck and A. Hilton. Surface capture for performance-based animation. *IEEE CGAA*, 2007.

[26] D. Terzopoulos, A. Witkin, and M. Kass. Symmetry-seeking models and 3d object reconstruction. *IJCV*, 1988.

[27] S. Valette, J. Chassery, and R. Prost. Generic remeshing of 3d triangular meshes with metric-dependent discrete voronoi diagrams. *IEEE TVCG*, 2008.

[28] M. Wand, B. Adams, M. Ovsjanikov, A. Berner, M. Bokeloh, P. Jenke, L. Guibas, H. Seidel, and A. Schilling. Efficient reconstruction of nonrigid shape and motion from real-time 3d scanner data. In *ACM Trans. Graph.*, 2009.

[29] K. Zhou, J. Huang, J. Snyder, X. Liu, H. Bao, B. Guo, and H. Shum. Large mesh deformation using the volumetric graph laplacian. 2005.