Nonrigid Surface Registration and Completion from RGBD Images*

Weipeng Xu[1,2], Mathieu Salzmann[2,3], Yongtian Wang[1], and Yue Liu[1]

[1] School of Optoelectronics, Beijing Institute of Technology (BIT), China
[2] NICTA, Canberra, Australia
[3] Australian National University (ANU), Australia

Abstract. Nonrigid surface registration is a challenging problem that suffers from many ambiguities. Existing methods typically assume the availability of full volumetric data, or require a global model of the surface of interest. In this paper, we introduce an approach to nonrigid registration that performs on relatively low-quality RGBD images and does not assume prior knowledge of the global surface shape. To this end, we model the surface as a collection of patches, and infer the patch deformations by performing inference in a graphical model. Our representation lets us fill in the holes in the input depth maps, thus essentially achieving surface completion. Our experimental evaluation demonstrates the effectiveness of our approach on several sequences, as well as its robustness to missing data and occlusions.

Keywords: Nonrigid registration, surface completion, RGBD images.

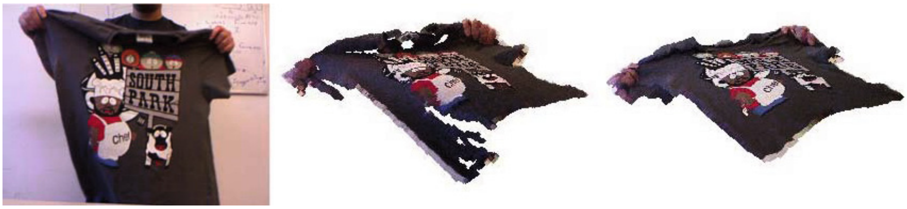
1 Introduction

Shape registration is a crucial process when handling sequences of 3D data. It allows us to progress from observing a set of unrelated point clouds to having a coherent 4D (space and time) shape representation, thus opening the way to a better understanding of the dynamic scene at hand. In the rigid scenario, many mature techniques are now available (e.g., [14,7]). Here, we focus on the challenging problem of nonrigid surface registration, which has many more degrees of freedom and typically suffers from many ambiguities.

In the nonrigid case, much of the existing literature has focused on registering volumetric data, where the point clouds represent (the surface of) a full 3D volume (e.g., [4,3,40,13,19,25,29]). Such volumetric data is typically acquired with multiple cameras placed around the scene of interest. Therefore, given sufficiently many cameras, it comes with relatively low noise and few holes in the point clouds. With the recent availability of low-cost RGBD sensors, 3D data can be acquired in much less constraining environments. However, this comes at the cost of **(i)** only having

* Electronic supplementary material -Supplementary material is available in the online version of this chapter at `http://dx.doi.org/10.1007/978-3-319-10605-2_5`. Videos can also be accessed at `http://www.springerimages.com/videos/978-3-319-10604-5`

D. Fleet et al. (Eds.): ECCV 2014, Part II, LNCS 8690, pp. 64–79, 2014.



(a) Input RGB image (b) Side view of the input data (c) Our registration result

Fig. 1. Nonrigid surface registration and completion: Given noisy RGBD images, such as the one depicted by (a) and (b), our approach computes a coherent shape across the images and lets us fill in the missing observations.

access to the 3D measurements of an inherently 2D surface instead of a 3D volume; and **(ii)** having to handle much noisier data with potentially many missing observations. While some techniques perform registration in this setting, with the exception of [10,11], they either acquire a model of the surface as a pre-processing step [15,16,6], or are designed for specific shapes, such as the human body [31].

In this paper, we introduce an approach to nonrigid surface registration from RGBD images that does not rely on a global surface model. To this end, we represent the observed surface as a collection of deformable patches. This makes our approach agnostic to the specific shape of the surface and thus applicable in more general scenarios. Furthermore, our patch-based representation gives us the means to fill in the gaps in the observations, and thus perform shape completion, as well as to account for newly visible surface parts in an online manner.

More specifically, we express nonrigid registration as an inference problem in a Conditional Markov Random Field (CRF) where each node represents a surface patch. The variables in our CRF are continuous and encode the deformations of the patches. This lets us formulate an energy function that comprises terms such as color constancy, coherence of the predicted 3D patches with the observed depth, surface smoothness and coherence across neighboring patches. The resulting energy is pairwise, which lets us employ a Fusion Moves strategy [18] to perform inference in our continuous CRF.

We evaluate our formulation on several sequences of deforming surfaces, such as the one depicted in Fig. 1. Our experiments evidence the benefits of our approach over model-free and model-based baselines in various scenarios including missing data and occlusions.

2 Related Work

In recent years, nonrigid registration has received a lot of attention. In particular, many methods have focused on the problem of registering, or matching, point clouds representing the surface of an entire 3D volume [4,3,40,13,19,25,29]. In this context, a patch-based representation of the surface was also employed in [4]. Note, however, that, in contrast to our deformable patches, the patches in [4] were assumed to move quasi-rigidly. More importantly, the above-mentioned techniques, as well as others designed for depth maps [20], ignore all appearance information of the object of interest. In [36,8,33], appearance was further included in the registration process. However, since the volumetric data was acquired with a multi-camera setup, the resulting techniques were able to exploit multiple views of the surface. In contrast, here, we make use of a low-cost RGBD sensor that only provides us with a single view of the surface, and whose 3D measurements are therefore not volumetric.

The measurements acquired with depth sensors are known to be noisy and to suffer from missing data. As a consequence, many techniques perform shape completion and denoising by fusing multiple depth images of the same scene [30,5]. This was further extended to incorporating appearance information in the fusion process [38,24]. While the previous methods rely on the availability of multiple depth maps of a rigid scene, depth super-resolution has also recently been investigated in the single image scenario. However, existing techniques require either having access to a large number of depth exemplars [21], or that the scene is redundant (i.e., has some repetitive patterns) [12].

Registering multiple depth maps to simultaneously perform denoising and build a model of a large scene has also been investigated [14,7]. While effective for rigid scenes, these techniques currently mostly treat nonrigid components as noise in the scene [17]. In [39], the nonrigid (articulated) case was explicitly handled. However, this was achieved by combining the information from multiple Kinects simultaneously acquiring images from different viewpoints.

Here, we aim to perform nonrigid registration from a single depth sensor. While some methods have been proposed to tackle this scenario, most of them rely on a global model of the observed shape. For instance, the technique of [31] is specifically designed for human pose estimation. While dealing with more general shapes, [6] relies on acquiring the global shape model from a separate RGBD sequence. In contrast, [15,16] build the shape model from the first frame of the sequence. However, this assumes that all parts of the surface are visible in this frame. To the best of our knowledge, [10,11] are the only RGBD-based methods that do not exploit knowledge of the shape at hand, and thus work in similar settings as our approach. Both methods, however, formulate the problem in terms of pixelwise flow, and thus are typically more sensitive to noise than model-based approaches.

Since RGBD sensors essentially provide us with 3D data corresponding to an inherently 2D surface (as opposed to full volumetric data), our work is also related to monocular nonrigid 3D reconstruction techniques [27]. These techniques can be categorized into template-based approaches [28,22,2] and nonrigid structure-from-motion methods [32,34,26,35,9]. Note that, in both classes, patch-based approaches were proposed [28,34,26]. As the name suggests, template-based methods make use of a reference model of the shape of interest. While nonrigid structure-from-motion techniques do not, they rely on tracking interest points across the sequence and are therefore sensitive to noise and occlusions. In contrast, since we exploit RGBD sensors that are now readily available in many scenarios, our approach can robustly track a surface throughout a sequence despite occlusions and missing data.

3 Nonrigid Surface Registration and Completion

We now present our approach to nonrigid surface registration. Ultimately, given a sequence of RGBD images depicting a deforming surface, our goal is to obtain a coherent surface representation across the entire sequence along with its deformations in each frame. We formulate this as nonrigid registration between pairs of neighboring frames, which lets us propagate information throughout the sequence. Note that, while we consider the video scenario, our registration approach could in principle be applied to any pair of images. In the remainder of this section, we first discuss our patch-based surface model and then describe our continuous CRF formulation of the nonrigid registration problem.

3.1 Nonrigid Patch-Based Surface Model

As mentioned earlier, we do not assume the availability of a global model of the surface of interest. This allows us to model surfaces of arbitrary shapes, and makes the design of deformation models easier. To this end, we represent the observed surface with a collection of overlapping, deformable patches.

Given an RGBD image of a nonrigid surface, we first compute a supervoxel segmentation of the observed data based on depth and color information [37]. Each supervoxel now corresponds to one patch in our representation. To parameterize the deformations of each patch, we make use of the linear deformation model of [28]. More specifically, each patch i is represented by a triangulated mesh with V vertices whose 3D coordinates are stored in the vector $\mathbf{m}_i \in \mathbb{R}^{3V}$. The deformation of this mesh can then be expressed as

$$\mathbf{m}_i = \mathbf{m}_i^0 + \Lambda \mathbf{c}_i \ , \tag{1}$$

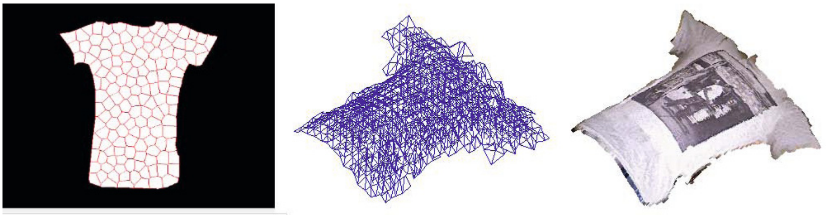
where $\mathbf{m}_i^0$ is the rest shape of the mesh, Λ is the $3V \times Q$ matrix of deformation modes, and $\mathbf{c}_i \in \mathbb{R}^Q$ is the vector of mode coefficients determining the deformation. As in [28], the deformation modes were obtained by applying Principal Component Analysis (PCA) to a set of randomly generated deformations of $\mathbf{m}_i^0$.

In this formalism, our goal now is to fit 3D mesh patches to the observed supervoxels. To this end, we first perform a rigid registration by aligning the centroid and the normal vector of patch i to that of its corresponding supervoxel, which yields $\mathbf{m}_i^0$. From this first registration, we estimate the barycentric coordinates $\mathbf{b}_{i,j} \in \mathbb{R}^3$ with respect to $\mathbf{m}_i^0$ of each observed point $\mathbf{x}_{i,j}$ belonging to supervoxel i. These barycentric coordinates are obtained by casting a ray from $\mathbf{x}_{i,j}$ in the direction of the supervoxel normal and intersecting it with $\mathbf{m}_i^0$. They are thus given with respect to 3 mesh vertices. Fitting $\mathbf{m}_i$ to the supervoxel points $\mathbf{x}_i = [\mathbf{x}_{i,1}^T, \cdots, \mathbf{x}_{i,N_i}^T]^T$ can be expressed as the optimization problem

$$\min_{\mathbf{c}_i} \left\| \mathbf{B}_i \left(\mathbf{m}_i^0 + \Lambda \mathbf{c}_i \right) - \mathbf{x}_i \right\|^2 + w_r \left\| \Sigma^{-1/2} \mathbf{c}_i \right\|^2 \ , \tag{2}$$

where $\mathbf{B}_i$ is the $3N_i \times 3V$ matrix grouping the barycentric coordinates of all the points, w_r is a regularization weight[1], and Σ is the diagonal $Q \times Q$ matrix

[1] The weight w_r was set to 3 in all our experiments.



(a) Supervoxel segmentation (b) Mesh fitting result (c) Point cloud

Fig. 2. Mesh fitting: Given the supervoxel segmentation (a), we fit a mesh patch to each supervoxel. In (c), we visualize the surface points computed using barycentric coordinates with respect to the mesh vertices.

containing the eigenvalues obtained when applying PCA to the training meshes to create the deformation modes. These eigenvalues encode the importance of each mode on the deformation. Thus the second term in the objective function simply regularizes the deformation coefficients. Note that this optimization problem is a least-squares problem. Therefore, its solution can be obtained in closed-form. In Fig. 2, we illustrate the process of fitting patches to observed supervoxels.

3.2 Nonrigid Registration as Inference in a CRF

In this paper, we address nonrigid registration as an inference problem in a CRF with continuous variables. By making use of the model introduced in Section 3.1, we define each node in our CRF as one surface patch. We employ the mode coefficients $\mathbf{c}_i$ as random variable for node i, which can be thought of as a Q-dimensional continuous label.

Given two images I_0 and I_1 and their corresponding depth maps D_0 and D_1, the joint distribution over the deformation coefficients of the N surface patches registering frame 0 to frame 1 can be expressed as

$$P(\mathbf{c}|I_0, I_1, D_0, D_1) = \frac{1}{Z(\mathbf{c})} \exp\left(-\sum_{i=1}^{N} \phi_i(\mathbf{c}_i|I_0, I_1, D_0, D_1) - w_p \sum_{(i,i')\in\mathcal{E}} \phi_{ii'}(\mathbf{c}_i, \mathbf{c}_{i'}) \right) ,$$

where $\phi_i(\cdot)$ is a unary potential function encoding the local evidence for a patch configuration, $\phi_{ii'}(\cdot,\cdot)$ is a pairwise potential function defined over the edges $\mathcal{E}$ of the CRF and accounting for the relation of two neighboring patches, w_p is a weight regulating the influence of both terms, and $Z(\cdot)$ is the partition function. Inference in the CRF is then performed by computing a MAP estimate of $\mathbf{c}$, which can be achieved by minimizing the corresponding energy

$$E(\mathbf{c}) = \sum_{i\in\mathcal{N}} \phi_i(\mathbf{c}_i|I_0, I_1, D_0, D_1) + w_p \sum_{(i,i')\in\mathcal{E}} \phi_{ii'}(\mathbf{c}_i, \mathbf{c}_{i'}) \ . \tag{3}$$

In the remainder of this section, we describe our unary and pairwise potentials, as well as the inference strategy used to obtain a MAP estimate.

Unary Potential. Our unary potential comprises three terms, and can thus be expressed as

$$\phi_i(\mathbf{c}_i) = w_c\phi_i^c(\mathbf{c}_i) + w_d\phi_i^d(\mathbf{c}_i) + w_s\phi_i^s(\mathbf{c}_i) \ , \tag{4}$$

where we omitted the explicit dependencies on the images and depth maps for ease of notation. These three terms encode color constancy (ϕ_i^c), depth coherence (ϕ_i^d) and patch smoothness (ϕ_i^s), respectively.

The color constancy term ϕ_i^c measures the intensity disagreement between corresponding pixels in I_0 and I_1. To this end, let us define the function $\Pi(\cdot)$, which projects a 3D point to its 2D image coordinates. The projection of a 3D point $\mathbf{x} = [x, y, z]^T$ can thus be written as

$$\mathbf{u} = \Pi(\mathbf{x}) = \begin{pmatrix} \frac{x}{z}f_x + c_x \\ \frac{y}{z}f_y + c_y \end{pmatrix} \ , \tag{5}$$

where f_x and f_y are the focal lengths of the sensor, and c_x and c_y define the optical center. This lets us write our color constancy term as

$$\phi_i^c(\mathbf{c}_i) = \left\| I_1 \left(\Pi\left(\mathbf{B}_i(\mathbf{m}_i^0 + \Lambda\mathbf{c}_i)\right)\right) - I_0\left(\Pi(\mathbf{B}_i\mathbf{m}_i^0)\right)\right\|_1 \ , \tag{6}$$

where $\mathbf{m}_i^0$ corresponds to the mesh configuration in frame 0 for patch i. Note that, to simplify notation, we assume that a vector input to the functions Π, I_0 and I_1 yields elementwise evaluations of the functions. Here, we utilize the ℓ_1 norm to increase the robustness of our method to outliers.

The depth coherence term ϕ_i^d penalizes depth disagreement between the deformed mesh patches and the target depth map D_1. By again making use of a robust ℓ_1 norm formulation, this term can be expressed as

$$\phi_i^d(\mathbf{c}_i) = \left\| D_1\left(\Pi\left(\mathbf{B}_i(\mathbf{m}_i^0 + \Lambda\mathbf{c}_i)\right)\right) - \left(\mathbf{B}_i(\mathbf{m}_i^0 + \Lambda\mathbf{c}_i)\right)_z\right\|_1 \ , \tag{7}$$

where $(\cdot)_z$ extracts the z-components of the input vector.

Finally, the smoothness term ϕ_i^s penalizes large deviations of the deformation coefficients $\mathbf{c}_i$ from the training data. Similarly as in Section 3.1, this term can be expressed as

$$\phi_i^s(\mathbf{c}_i) = \left\| \Sigma^{-1/2}\mathbf{c}_i \right\|^2 \ , \tag{8}$$

and can be thought of as assuming a Q-dimensional zero-mean Gaussian prior on $\mathbf{c}_i$ with (diagonal) covariance Σ.

Pairwise Potential. Our unary potentials try to fit each individual mesh patch to the data observed in frame 1. Of course, this can be subject to ambiguities, since parts of the surface may be poorly-textured. This in turn would yield to inconsistencies between the different patches. To prevent these inconsistencies, we introduce a pairwise potential defined on adjacent mesh patches, i.e., patches that overlap when fitting the meshes to the supervoxels.

More specifically, let us denote by $\hat{\mathbf{x}}_{i,j}$ the 3D location of point j predicted by patch i. This prediction can be written as

$$\hat{\mathbf{x}}_{i,j} = \mathbf{B}_{i,j}(\mathbf{m}_i^0 + \Lambda\mathbf{c}_i) \ , \tag{9}$$

where $\mathbf{B}_{i,j}$ is the $3 \times 3V$ submatrix of $\mathbf{B}_i$ corresponding to point j. Similarly, if point j also belongs to patch i', we have the prediction

$$\hat{\mathbf{x}}_{i',j} = \mathbf{B}_{i',j}(\mathbf{m}^0_{i'} + \Lambda\mathbf{c}_{i'}) \ . \tag{10}$$

To enforce coherence across adjacent patches, our pairwise potential therefore encourages these two predictions to agree. By summing over the $N_{i,i'}$ points in the overlap between patch i and patch i', this can be written as

$$\phi_{ii'}(\mathbf{c}_i, \mathbf{c}_{i'}) = \sum_{j=1}^{N_{i,i'}} \left\| \mathbf{B}_{i,j}(\mathbf{m}^0_i + \Lambda\mathbf{c}_i) - \mathbf{B}_{i',j}(\mathbf{m}^0_{i'} + \Lambda\mathbf{c}_{i'}) \right\|^2 \ . \tag{11}$$

Together, our unary and pairwise potentials define an energy that encourages the patches to fit the data and form a coherent surface. We now turn to the problem of minimizing this energy, i.e., performing inference in our CRF.

Inference. One of the challenges of performing inference in our graphical model is that its variables are continuous. While many inference methods have been proposed for the discrete case, the literature on continuous inference remains limited. Here, we perform inference using the Fusion Moves strategy introduced in [18]. Fusion Moves have proven effective to minimize nonlinear and non-convex objective functions in different contexts, such as optical flow estimation [18] and monocular nonrigid surface reconstruction [35]. In our framework, they are particularly better suited than gradient-based techniques to handle non-smooth robust error terms, such as ϕ^c_i and ϕ^d_i. Fusion Moves perform inference by iteratively merging a set of two proposals per node in the graph, which can be expressed as a binary labeling problem. While global convergence cannot be ensured, Fusion Moves guarantee that the energy decreases at each iteration.

As mentioned earlier, our deformation model was computed via PCA, and thus assumes a Gaussian distribution of the deformation coefficients. To generate proposals, we therefore make use of this assumption, and draw random samples for each patch according to

$$\mathbf{c}_i \sim \mathcal{N}(\mathbf{0}, \varepsilon\Sigma) \ , \tag{12}$$

where we exploit the covariance matrix of the coefficients Σ modulated by a constant scalar ε that encodes the degree of deformation expected between frame 0 and frame 1[2]. At each iteration, we draw one such sample per patch and make use of the current best solution as second proposal in the Fusion Moves. We iterate until the change of global energy is smaller than a predefined tolerance within ten iterations.

To tackle the video scenario, which we are most interested in, we use the method described in Section 3.1 to obtain the initial surface patches in the first frame of the video. We then track the deformations of these patches throughout the entire sequence by making use of the solution at time t as initialization

[2] We used $\varepsilon = 0.05$ in all our experiments.

for the Fusion Moves inference at time $t + 1$. As a consequence, we obtain a coherent surface model across the video sequence. In the presence of occlusions and missing data, the unobserved parts of the surface can then be inferred from the deformed mesh patches, thus effectively performing surface completion. Note that, while we do not model occlusions explicitly, our method has proven robust to moderate occlusions. For larger ones, a strategy such as the one used in [16] could easily be included in our framework. As will be evidenced by our experiments, our inference strategy makes our approach robust to error accumulation that can typically occur in a tracking scenario such as considered here.

3.3 Incorporating New Patches

One of the benefits of relying on patches instead of a global model is that we can easily incorporate new patches in an online manner, as previously-hidden parts of the surface become visible. To this end, we adopt the following strategy. After registering the previous frame to the current one, we search for parts in the point cloud that are not explained by the current patches. These parts correspond to newly visible surface areas. We then compute superpixels [1] in the input image corresponding to these areas only, and add a new patch for each sufficiently large superpixel (i.e., of similar size as the original patches). These patches are then registered to the RGBD image following the procedure of Section 3.1, and thus incorporated in our surface representation for the following frame.

4 Experimental Results

In this section, we study the effectiveness of our approach in different scenarios. In particular, we provide quantitative and qualitative results on two publicly available RGBD sequences[3] depicting dynamically deforming objects (a sheet of paper and a T-shirt) captured using a single Kinect. We also evaluate the robustness of our approach to missing data and occlusions, and illustrate its use to add patches online. In all our experiments, both the color images and the depth maps were acquired at the resolution of 640×480. We used square mesh patches with 5×5 vertices and retained $Q = 20$ deformation modes. The weights in Eqs. 3 and 4 were set to $w_p = 2500$, $w_c = 30$, $w_d = 25$, and $w_s = 10000$ for all sequences, which corresponds to making all the terms in the energy of similar magnitudes. The videos of our results are provided as supplementary material.

We compare our results with those obtained with two methods that, as ours, do not require a global model of the surface: (i) The Coherent Point Drift (CPD) algorithm of [23][4], which does not exploit RGB intensities; and (ii) RGBD-Flow [11][5], which extends state-of-the-art optical flow techniques to RGBD images. Furthermore, we also compare our results with a baseline obtained by replacing our local patches with the global NURBS representation of [16], in

[3] Publicly available at `http://cvlab.epfl.ch/data/dsr`

[4] Code available at `https://sites.google.com/site/myronenko/research/cpd`

[5] Code available at `http://homes.cs.washington.edu/~eherbst/useful-code/`

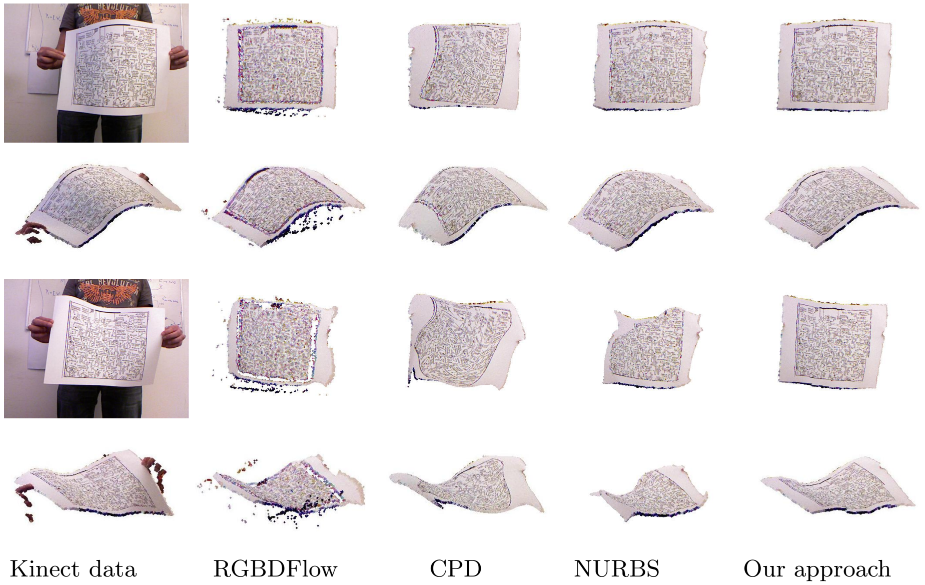


Fig. 3. Registration results for 2 frames of the paper sequence. We show the front view and side view of our results and of the baselines. Note that RGBDFlow introduces some noise in the results, while CPD and the NURBS model do not respect the texture of the surface. Our approach yields smooth and accurate surfaces.

conjunction with their regularizer and optimization technique (CMA-ES). For all baselines, we follow a similar strategy as with our approach to propagate information across the video sequence.

4.1 Comparison with the Baselines

In Figs. 3 and 4, we provide some examples of the registration results obtained with our approach and with the baselines on the paper and T-shirt sequences. To be able to better evaluate the registration quality, we warped the first RGBD image according to the registration results, and display the result of this warping procedure as colored point clouds seen from a different viewpoint.

From these examples, we can see that, while reasonably accurate in the middle of the surface, RGBDFlow yields very noisy point locations at the boundaries. In contrast, CPD and the NURBS model yield fairly smooth surfaces. However, the artifacts that can be observed on the warped texture of the surface suggest that registration is not very accurate. Our method simultaneously yields smooth results and accurate warped texture, thus indicating accurate registration.

The benefits of our approach are further evidenced by our quantitative results. In Fig. 5, we report depth errors for all methods, computed as the mean absolute depth difference between depth maps generated from the results of the different methods and the original Kinect depth maps. Note that the error of RGBDFlow

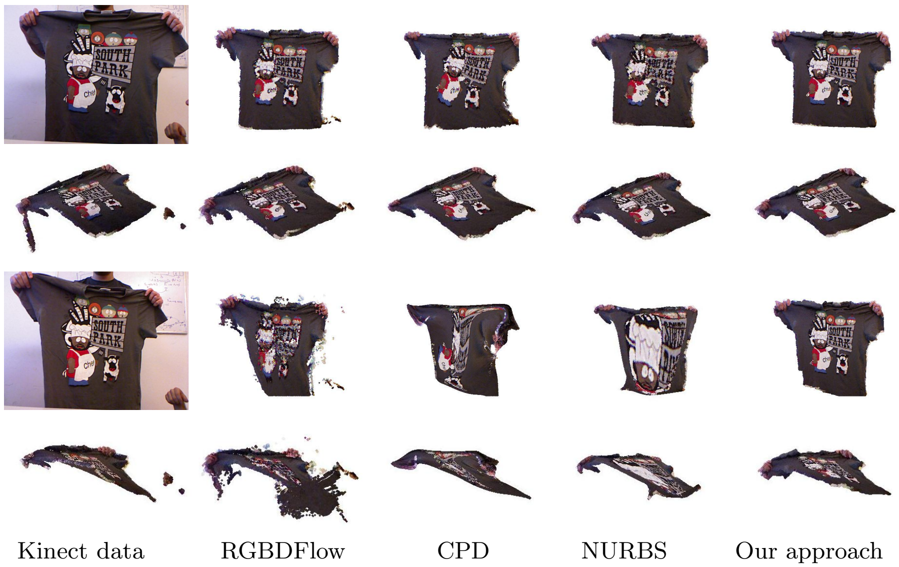


Fig. 4. Registration results for 2 frames of the T-shirt sequence. Note again that our approach yields smoother results than RGBDFlow and less distorted warped textures than CPD and the NURBS model, thus indicating more accurate registration.

increases over time, thus suggesting error accumulation. In contrast, CPD, the NURBS model and our approach yield lower errors and are more stable over time. Importantly, even when our accuracy drops for one frame (e.g., frame 45 in the T-shirt sequence), our Fusion Moves inference strategy, which performs more global optimization, is able to recover and get back to accurate registration in later frames. While it may seem that CPD, the NURBS model and our approach perform similarly, the front views of Figs. 3 and 4 suggest that our registration results are more accurate. To quantify this, we employed the following strategy: We manually selected 15 interest points spread over the surface every 20 frames of the paper and T-shirt sequences. The shapes predicted by our algorithm and the baselines give us estimates of the 3D locations of these points, which we can then compare against the manually selected locations. This gives us a better notion of registration error than the previous depth error, which does not take point correspondence into account and thus does not penalize points sliding on the surface. As evidenced by Fig. 6, our approach yields more accurate registration than the baselines.

On a laptop with a four-core CPU, 8GB memory and a Geforce 765M graphics card, the runtimes for the 4 methods are of the order of 1 min per frame for RGBDFlow, 10 min per frame for CPD, 10 min per frame for the NURBS model and 3 min per frame for our approach. Note that RGBDFlow benefits from a GPU implementation. In contrast, the current implementation of our approach only runs on the CPU.

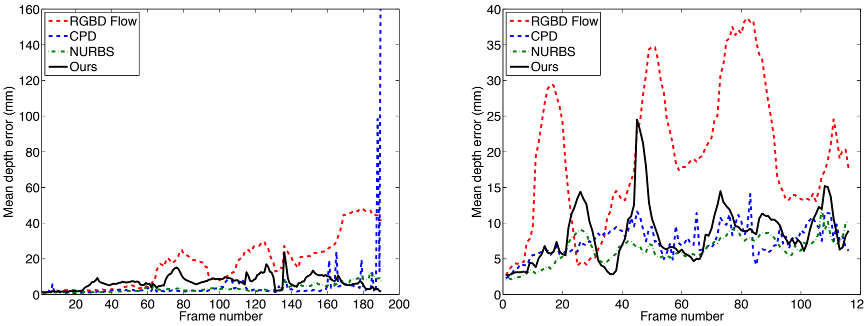


Fig. 5. Mean depth error: (Left) Paper sequence. (Right) T-shirt sequence.

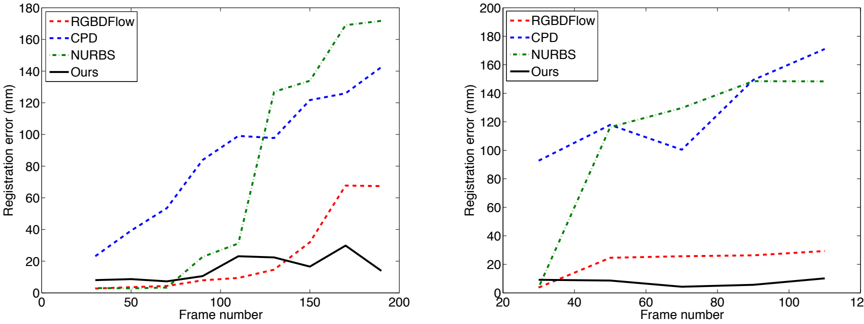


Fig. 6. Registration error: (Left) Paper sequence. (Right) T-shirt sequence.

4.2 Missing Data and Occlusions

In this section, we evaluate the robustness of our algorithm to missing data and partial occlusions. For the missing data case, we first made use of the paper sequence, and synthetically removed some amount of depth observations. To this end, and to model the fact that the noise in the depth sensor is often structured, we randomly removed square areas of 20 × 20 pixels from the depth maps of 10 consecutive frames in the sequence. We repeated this operation for different numbers of removed areas, corresponding to different percentages of missing data. Fig. 7 depicts the mean depth errors of the baselines and of our approach on the last of the 10 frames as a function of the percentage of missing points. Note that, while the errors of RGBDFlow and of the NURBS model tend to increase with the percentage of missing data, our errors remain stable. While the errors of CPD, and in some cases of the NURBS model, appear to be very low, they should again be considered jointly with the front view of the results (see Fig. 9) which suggests worse registration accuracy than for our approach.

To illustrate the fact that our approach is also robust to the missing data problem in a real scenario, we selected the images in the T-shirt sequence that

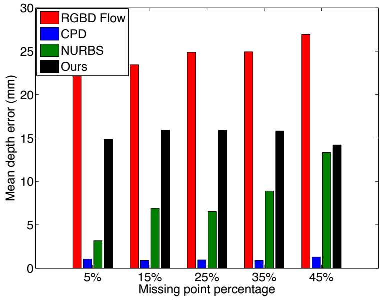


Fig. 7. Robustness to missing data: Mean depth error of the baselines and of our approach as a function of the percentage of missing points. The data was generated by randomly removing areas from some frames of the paper sequence

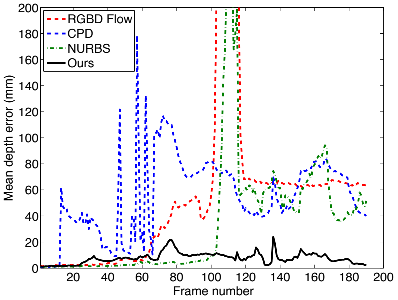
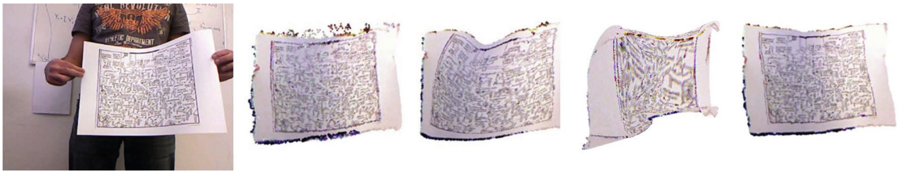


Fig. 8. Robustness to occlusion: Mean depth error of the baselines and of our approach as a function of the frame number. The data was generated by adding an occluder to the paper sequence



Kinect data RGBDFlow CPD NURBS Our approach

Fig. 9. Robustness to missing data: Front view of the registration results for the largest amount of missing data in Fig. 7

most suffer from this problem. These results are depicted in Figs. 1 and 10. Note that we can accurately fill in the holes in the original Kinect data.

To evaluate the robustness of our approach to partial occlusion, we employed the paper sequence and synthetically augmented it with an occluder passing in front of the surface, as shown in the first column of Fig. 11. Note that the resulting depth maps contained the depth of the occluder, thus truly mimicking occlusion. Fig. 11 compares our results on this occluded sequence with the results of the baselines. While the latter are strongly affected by the presence of the occluder, our approach performs essentially the same as without any occlusion. This is further evidenced by the quantitative errors provided in Fig. 8. Note that our results can also be used to fill in the gap in the observed data.

4.3 Incorporating New Patches

To illustrate the ability of our approach to incorporate new patches in an online manner, we employed the paper sequence, and augmented the data with a synthetic occluder that hides roughly half of the surface in the first frame. This

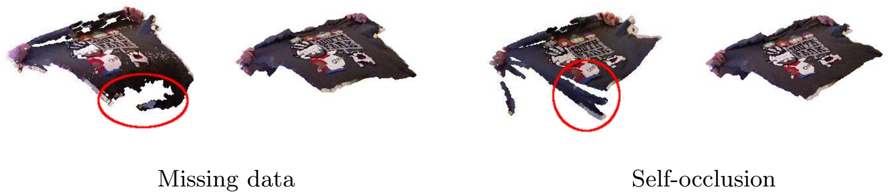


Fig. 10. Missing data and self occlusion: Results on some of the most noisy frames of the T-shirt sequence

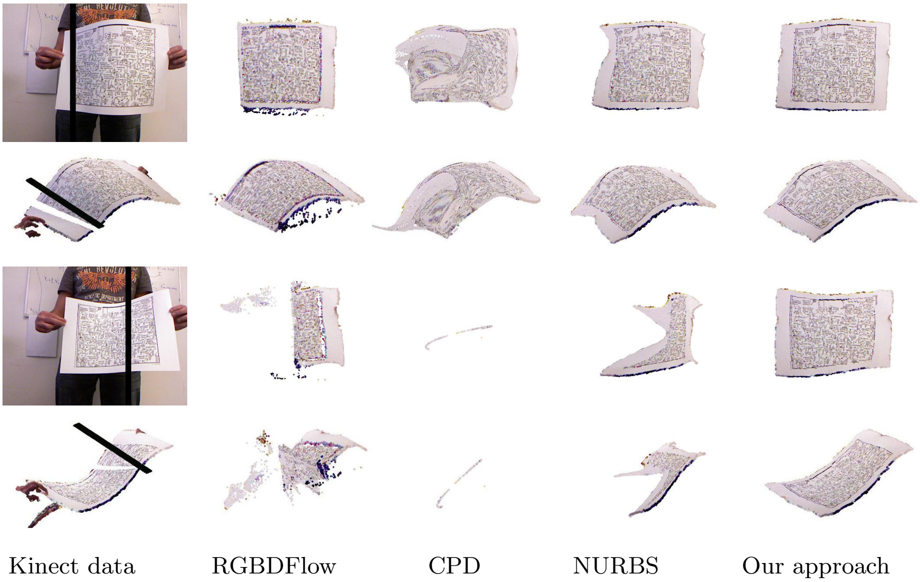


Fig. 11. Robustness to occlusions: While the baselines all suffer from the presence of an occluder in the sequence, our approach remains virtually unaffected

occluder was then progressively removed throughout the sequence. As shown in Fig. 12, our approach lets us augment the surface with new patches as parts of the object that were previously hidden become visible. Note that this would not be possible with a global model acquired in the first frame of the sequence.

4.4 Video Editing

Finally, we illustrate the use of our approach as a tool to achieve video editing. To this end, we re-textured the data in the first frame of the video and then made use of our registration results to transfer the new texture across the entire sequence. Fig. 13 depicts a few frames of the result of this process applied to the paper sequence. Note that, thanks to our registration accuracy, the resulting images look realistic, up to shading effects which are not modeled here.

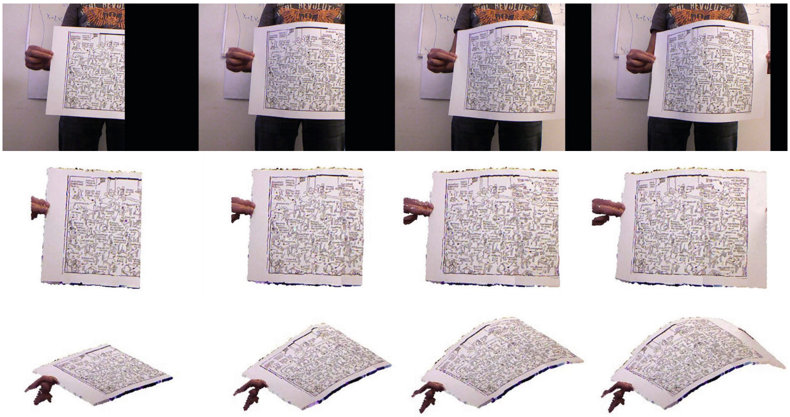


Fig. 12. Incorporating Patches: Our approach lets us incorporate new patches as previously hidden parts of the surface become visible

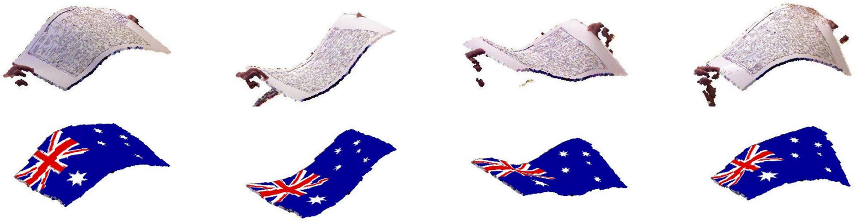


Fig. 13. Video editing: Our registration results let us re-texture the surface in the paper sequence to create realistic RGBD images of a different object

5 Conclusion

We have introduced an approach to performing nonrigid surface registration and completion from a low-cost RGBD sensor. Thanks to our patch-based representation, our approach can be applied to surfaces of arbitrary shapes. Our experimental evaluation has demonstrated the effectiveness of our method in various scenarios. In the future, we plan to push our technique in the direction of augmented reality, as well as extend it to modeling more general 3D scenes.

Acknowledgement: This work was supported in part by National Program on Key Basic Research Project of China (973 Program) 2013CB328805, National High-tech R&D Program of China (863 Program) 2012AA013802 and National Natural Science Foundation of China (NSFC) 61370134. The authors would also acknowledge the financial support from China Scholarship Council. NICTA is funded by the Australian Government as represented by the

Department of Broadband, Communications and the Digital Economy, as well as by the Australian Research Council through the ICT Centre of Excellence program.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(11), 2274–2282 (2012)
2. Bartoli, A., Pizarro, D., Collins, T.: A robust analytical solution to isometric shape-from-template with focal length calibration. In: ICCV (2013)
3. Bronstein, A.M., Bronstein, M.M., Kimmel, R., Mahmoudi, M., Sapiro, G.: A Gromov-Hausdorff Framework with Diffusion Geometry for Topologically-Robust Non-Rigid Shape Matching. IJCV (2010)
4. Cagniart, C., Boyer, E., Ilic, S.: Free-form mesh tracking: a patch-based approach. In: CVPR (2010)
5. Cui, Y., Schuon, S., Chan, D., Thrun, S., Theobalt, C.: 3d shape scanning with a time-of-flight camera. In: CVPR (2010)
6. Dou, M., Fuchs, H., Frahm, J.M.: Scanning and tracking dynamic objects with commodity depth cameras. In: ISMAR (2013)
7. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., Burgard, W.: An evaluation of the rgb-d slam system. In: ICRA (2012)
8. Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H.P.: Motion capture using joint skeleton tracking and surface estimation. In: CVPR (2009)
9. Garg, R., Roussos, A., Agapito, L.: Dense Variational Reconstruction of Non-Rigid Surfaces from Monocular Video. In: CVPR (2013)
10. Hadfield, S., Bowden, R.: Kinecting the dots: Particle based scene flow from depth sensors. In: ICCV (2011)
11. Herbst, E., Ren, X., Fox, D.: Rgb-d flow: Dense 3-d motion estimation using color and depth. In: ICRA (2013)
12. Hornacek, M., Rhemann, C., Gelautz, M., Rother, C.: Depth super resolution by rigid body self-similarity in 3d. In: CVPR (2013)
13. Huang, P., Budd, C., Hilton, A.: Global temporal registration of multiple non-rigid surface sequences. In: CVPR (2011)
14. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In: UIST (2011)
15. Jordt, A., Koch, R.: Fast tracking of deformable objects in depth and colour video. In: BMVC (2011)
16. Jordt, A., Koch, R.: Direct model-based tracking of 3d object deformations in depth and color video. IJCV (2013)
17. Keller, M., Lefloch, D., Lambers, M., Izadi, S., Weyrich, T., Kolb, A.: Real-time 3d reconstruction in dynamic scenes using point-based fusion. In: 3DV (2013)
18. Lempitsky, V., Rother, C., Roth, S., Blake, A.: Fusion moves for markov random field optimization. PAMI (2010)
19. Letouzey, A., Boyer, E.: Progressive shape models. In: CVPR (2012)
20. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. In: SGP (2008)

21. Mac Aodha, O., Campbell, N.D.F., Nair, A., Brostow, G.J.: Patch based synthesis for single depth image super-resolution. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 71–84. Springer, Heidelberg (2012)
22. Malti, A., Hartley, R., Bartoli, A., Kim, J.H.: Monocular template-based 3d reconstruction of extensible surfaces with local linear elasticity. In: CVPR (2013)
23. Myronenko, A., Song, X.: Point set registration: Coherent point drift. PAMI (2010)
24. Park, J., Kim, H., Tai, Y.W., Brown, M.S., Kweon, I.: High quality depth map upsampling for 3d-tof cameras. In: ICCV (2011)
25. Rouhani, M., Sappa, A.D.: Non-rigid shape registration: A single linear least squares framework. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VII. LNCS, vol. 7578, pp. 264–277. Springer, Heidelberg (2012)
26. Russell, C., Fayad, J., Agapito, L.: Energy Based Multiple Model Fitting for NRSFM. In: CVPR (2011)
27. Salzmann, M., Fua, P.: Deformable Surface 3D Reconstruction from Monocular Images. Morgan Kaufmann (2010)
28. Salzmann, M., Fua, P.: Linear Local Deformation Models for Monocular Reconstruction of Deformable Surfaces. PAMI (2011)
29. Santa, Z., Kato, Z.: Correspondence-less non-rigid registration of triangular surface meshes. In: CVPR (2013)
30. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: CVPR (2009)
31. Taylor, J., Shotton, J., Sharp, T., Fitzgibbon, A.: The vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation. In: CVPR (2012)
32. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid Structure-From-Motion: Estimating Shape and Motion with Hierarchical Priors. PAMI (2008)
33. Ulusoy, A.O., Biris, O., Mundy, J.L.: Dynamic probabilistic volumetric models. In: ICCV (2013)
34. Varol, A., Salzmann, M., Tola, E., Fua, P.: Template-Free Monocular Reconstruction of Deformable Surfaces. In: ICCV (2009)
35. Vicente, S., Agapito, L.: Soft inextensibility constraints for template-free non-rigid reconstruction. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 426–440. Springer, Heidelberg (2012)
36. Vlasic, D., Baran, I., Matusik, W., Popovic, J.: Articulated mesh animation from multi-view silhouettes. In: SIGGRAPH (2008)
37. Weikersdorfer, D., Gossow, D., Beetz, M.: Depth-adaptive superpixels. In: ICPR (2012)
38. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: CVPR (2007)
39. Ye, G., Liu, Y., Hasler, N., Ji, X., Dai, Q., Theobalt, C.: Performance capture of interacting characters with handheld kinects. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 828–841. Springer, Heidelberg (2012)
40. Zeng, Y., Wang, C., Wang, Y., Gu, X., Samaras, D., Paragios, N.: Dense non-rigid surface registration using high-order graph matching. In: CVPR (2010)