Markerless Real-Time Garment Retexturing from Monocular 3D Reconstruction

George A Cushen, Mark S Nixon

University of Southampton, UK
{gc505, msn}@ecs.soton.ac.uk

Abstract—We present a fusion of augmented reality (AR) and virtual try on (VTO) that incorporates sparse 3D point recovery by exploiting distance constraints based on 2D point correspondences between a deformed texture in monocular video and a reference texture which is derived from the start of the sequence by face detection aided segmentation. A hierarchical and multi-resolution surface reconstruction approach is proposed, employing thin-plate splines, cloth modeling, and patch tessellation. Our method attempts to accurately recover a rectangular surface from a deformed arbitrarily shaped texture. We also propose a hue-based method for segmenting garment cloth and printed texture. The methods are demonstrated in an AR framework for real-time visualization of a virtual garment worn in a real scene. Real-time AR cloth retexturing from monocular vision is a state-of-the-art field. Previous work lacks realism and accuracy, only recovering the 2D cloth layout and lacks robustness, requiring a special T-shirt color and simple texture along with lab hardware. Our approach alleviates these limitations. We design a practical approach which considers a typical consumer environment with a mid-range PC and webcam. Our results are convincing and photorealistic with robustness to arbitrary T-shirts, subjects, and backgrounds. Future work will focus on extending our global model and quantitative analysis.

I. Introduction

Augmented Reality (AR) is the concept of adding virtual elements to real world scenes. Virtual Try On (VTO) is the concept of allowing a user to try on garments virtually to check the fit and look of garments on a textured body model of themselves. In contrast to AR, the VTO concept is traditionally a virtual reality environment in which the real world is entirely replaced with a simulated one where the scene, user and clothing are modeled and computer generated.

Virtual try on is revolutionizing the clothing retail and fashion industries [1]. However, existing VTO methods can be complex to implement and are computationally intensive requiring an offline setup with manual user interaction for 3D body modeling and garment design. It is also challenging to model and texture a 3D representation of the user which is both photorealistic and anthropometrically accurate given a practical non-invasive consumer driven input such as a single 2D photo. *Augmented Try On (ATO)* can be achieved by fusing the concept of AR with VTO, allowing for systems which are more real-time and photorealistic to be designed with a focus on visualization of the look/style of the retextured garment worn by the user rather than on checking the fit. This fusion may be particularly well suited for visualization of printed textures on common garments whose designs do not vary as much as other types of garment and for which users will likely already know their size.

In this work, we describe a *practical* and *photorealistic* method for reconstruction of cloth geometry and retexturing for augmented reality in a non-lab environment.

Our contributions are:

- a hierarchical multi-resolution method employing thin-plate splines, cloth modeling, and patch tessellation for detailed and dynamic 3D geometric reconstruction of highly deformable surfaces (such as cloth) from a set of sparse 3D points in real-time. A local thin-plate spline model recovers a continuous rectangular surface mesh from a limited set of tracked features on a partially textured cloth. A global cloth model then attempts to (a) increase the accuracy of this reconstruction at regions with less texture and fewer features such as the untextured cloth around arbitrarily shaped textures, and (b) restore cloth dynamics. Patch tessellation is employed to subdivide the mesh, increasing surface smoothness and improving realism at regions of high curvature.

- a real-time method for reasonably accurate segmentation of the cloth of a plain-colored garment and arbitrary printed region(s) within it which is robust to variation in T-shirt shape and color, print shape and color, subject, and background. The approach is automated and without requirement of user interaction.

- we demonstrate the methods in a framework for garment retexturing for the purpose of real-time augmented reality in video. It is designed to be practical for a consumer and unlike most related work which utilizes high-end cameras and special lighting, no expensive high-end hardware is required or used for obtaining results.

II. Previous Work

Our methods are primarily related to work from the following three areas of research: (1) geometric recovery of flexible surfaces, (2) garment segmentation, and (3) real-time augmented reality. This section describes the most relevant work from each of these categories.

Recovery of Flexible Surfaces: The *2D* recovery of a non-rigid surface from *monocular vision* can generally be considered to be solved [2], [3]. Recently, [4] and [5] have extended the work on this topic for the purpose of AR applications. Bradley et al. use markers printed onto a white T-shirt for 2D cloth recovery and optical flow is used to track

the marker positions over a video sequence for retexturing the markered region with a new image. Hilsmann and Eisert extend this optical flow approach with self-occlusion handling and consider a specific T-shirt with a rectangular texture rather than markers.

Multiple view geometry is a very popular field in Computer Vision for solving problems such as 3D shape reconstruction. A survey is presented in [6]. Given sufficient views, this process is over-constrained for a moderately deformable surface. However, for the case of a highly non-rigid surface, such as cloth, the cloth can exhibit self-occlusions which make simple point correspondence techniques fail. Recently, [7] addressed this, presenting a markerless approach.

Recovering the *3D* geometric layout of deformable self-occluding surfaces from *monocular vision* remains an open and challenging problem due to the fact that it is severely under-constrained. The problem can be overcome to some extent by introducing deformation models which are either physically-based [8] or learned from training data [9]. Good initialization is important but very difficult to achieve in these methods because the image based objective function, that is minimized to obtain model parameters, generally has many local minima. Recent work [10], [11], [12] addresses this issue by proposing constraints of a reference image in which the shape is known, and pixel correspondences between the input and reference images which are also known. These methods assume that the surface is inextensible.

Garment Segmentation: The field of garment segmentation is relatively unexplored, but has recently received much research attention for the purpose of using clothing as a cue to help identify a person from various images captured over the same day. One of the most popular approaches involves graph cuts [13], [14]. Although this approach is robust to a diverse range of garments and is efficient, it can suffer in accuracy, especially in cases of occlusions and difficult poses, and is not suitable for real-time augmented reality. Additionally, we are interested in not only segmenting cloth, but also printed texture regions on the cloth. The aforementioned papers are not designed to handle this. Color-based segmentation is another approach. In [5], an RGB-colorspace approach is briefly described for segmenting T-shirt cloth and texture in a manner which to some degree removes these drawbacks. However, their method requires major *a prioris* such as cloth color and a simple rectangular texture, and they only show subjective visual results for a specific T-shirt.

Augmented Reality: A survey is presented in [15]. The previously described work in [4] and [5] can be considered the closest to our AR framework.

III. OVERVIEW

Figure 1 depicts an overview of our framework. After initialization has taken place at the start of a real-time monocular video sequence, this loop is iterated for each remaining frame. The core contributions are highlighted and are described in the following sections.
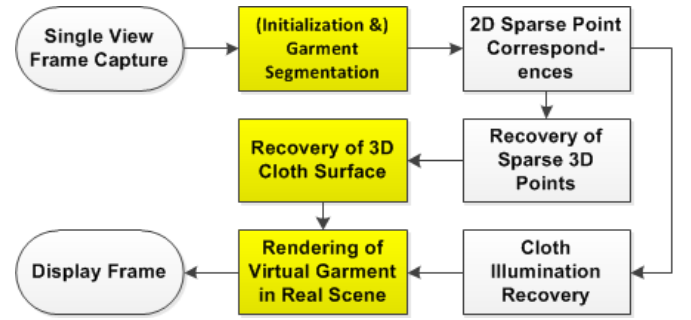

Fig. 1: Framework Overview.

Fig. 2: Our hue based segmentation: (a) - (f) T-shirt cloth segmentation; (e) additionally shows sampling of the T-shirt color (yellow boxes) using a simple torso model initialized automatically by face detection (red box); (g) - (h) texture segmentation; (i) texture template used for point correspondences; (j) T-shirt color recovered from (e); illumination recovery (k) before; and (l) after.

For image acquisition, we use a single Microsoft LifeCam Cinema webcam at a resolution of 480×640. Our reconstruction method requires geometric calibration of the camera. We use the popular chessboard technique implemented in the OpenCV library in order to calibrate the intrinsic and extrinsic camera parameters. Lighting also has to be considered. Since we focus on a practical application for demonstration, we choose to primarily work with images captured under uncontrolled domestic lighting.

During pre-processing, each frame captured by the webcam is converted from RGB to HSV colorspace and the corresponding illumination channel is normalized. This simple online process helps alleviate, to some extent, the non-uniform effects of uncontrolled domestic lighting.

IV. GARMENT SEGMENTATION

Segmentation of the textured garment is required in order to find the flexible textured surface which we wish to reconstruct and also for allowing the garment to be recolored during rendering. For demonstration, we consider the segmentation of loose T-shirts since they exhibit highly non-rigid properties and often feature a printed texture, which is required for our reconstruction and retexturing methods.

The segmentation of clothing worn on a subject is challenging due to the wide diversity of clothing designs, and the complexity of scene lighting, dynamic backgrounds and self/third-party occlusions. The main previous work has major aforementioned *a prioris* such as T-shirt color and a simple rectangular texture, and they only show subjective visual results for a specific T-shirt.

The initialization procedure for our segmentation approach consists of two simple steps for the user: (a) garment color recovery: stand approximately $1m$ or so from the camera so that the full face and torso are in view; and (b) garment texture recovery: attempt to hold the T-shirt so that the texture is roughly planar and perpendicular to the camera's optical axis. Our approach has been designed to be intuitive and require minimum keyboard/mouse interaction.

Cloth Color Recovery: We design a hue-based technique for segmentation of the T-shirt cloth as this approach is suitable for real-time applications whilst also providing robustness to different T-shirt shapes and sizes. Unlike recent AR retexturing work, we do not require a known T-shirt color. We remove this limitation by designing a method to automatically recover the color. We employ fast face detection [16] at the start of the video sequence. OpenCV includes a popular and reliable trained classifier cascade for frontal face, so we employ this existing cascade rather than training the classifier ourselves. We limit the region of interest for object detection to the top half of the image to increase efficiency.

Once the closest and largest head is found (i.e. there is robustness to spectators in the background), the normalized colors of two patches each situated beneath the collar bones are sampled. The colors within this temporary array of patches which have stable hues (no illuminations extremities or low saturations) are saved into vector **C**. The previous steps are repeated across new frames until stable colors from three sets of patches have been collected. These repetitions add robustness to noise, which is particularly important in the case of a noisy webcam. Finally, **C** is averaged to estimate the plain T-shirt color.

The size and position of the patches are dynamically adjusted based on the bounding box size of the detected face. Hence there is robustness to different subjects and setups. For example, if the user steps backwards, the patch size and relative distance from the face will decrease in order to attempt to maintain the patches within the torso and below the collar bones. The positions of the patches relative to the face are chosen due to the fact they are generally well lit from ceiling lights and are above the region which screen-printed textures often occupy. The max bounds of this potentially textured region were computed from a random sample of 25 people wearing plain T-shirts containing a highly textured design. The random sample was taken from the Soton [17] dataset.

Cloth Segmentation: A T-shirt segmentation mask is required for finding the unknown arbitrary printed garment texture and for recoloring the garment during the rendering stage.

Now that the T-shirt color is known, we describe the core part of the segmentation routine. First, the inputs are prepared. Rather than employing the single hue obtained in the previous section for guiding the segmentation, we increase robustness by considering the entire set **C** of pixels with stable hues. A histogram is computed for the hue channel of the pixel vector **C**. We empirically choose to quantize to 16 bins as this provides a good compromise between under-segmentation (due to variation in T-shirt hue caused by lighting) and over-segmentation (due to clothes with similar hues which are in direct contact with the T-shirt). Quantization reduces the computational and space complexity for analysis, clustering similar color values together. The histogram is then normalized to the discrete range of image intensities.

We filter out pixels with unstable hues (colors) in the current frame. Unstable hues are those pixels which feature illumination extremes and/or low saturations. The remaining hue channel of the current frame is back-projected. Back-projection is a histogram operation that associates the pixel values in the frame with the value of the corresponding histogram bin, generating a probability distribution image where the value of each pixel characterizes the likelihood of it belonging to the T-shirt (i.e. the histogram computed for **C**). Let $b: \mathbf{R}^2 \to \{1 \ldots 16\}$ be the back-projection function which maps the pixel at location (x, y) to the histogram bin index $b(x, y)$. The resulting probability image is thresholded to create a binary image from the high probability pixels. We fill small holes which can be caused by optic noise (especially in the case of a webcam). Given the next step, the main purpose here is to improve the T-shirt edges. The approach by [18] is employed to extract the largest contour in the binary image for the purpose of adding robustness to objects in the scene which have a similar hue as the T-shirt. The T-shirt segmentation mask is defined by filling this contour.

Texture Segmentation: Garment texture segmentation is critical for finding the template texture and deformed texture in the video frames in order to establish point correspondences, and for the purpose of illumination recovery across the texture during the rendering stage. During the final stage of initialization, *and for initialization only*, we assume that the printed region of the cloth's surface is planar. In practice, the surface region of interest resting on a flat chest can achieve a generally planar result, or more ideally a thin planar object can be pressed up behind the surface either with the garment worn or taken off.

First we find the largest holes above an area threshold within an intermediate T-shirt segmentation result prior to filling the mask. We decide that any T-shirt hue occurring
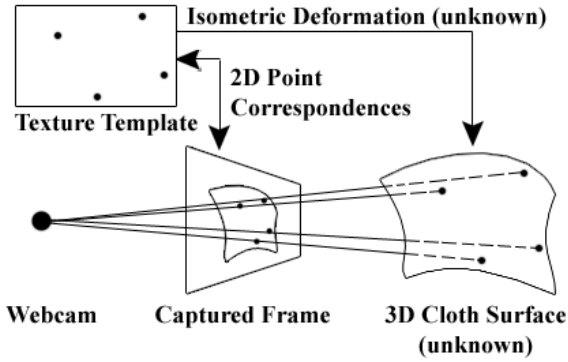
Fig. 3: Cloth layout recovery.

within the texture boundaries is not part of the texture, adding robustness for arbitrary texture designs where the contour(s) of the texture are not required to be entirely filled with texture. Based on the aforementioned random sample of people, we find this is usually a valid assumption and has the advantages of decreasing the computation time required for illumination recovery later on as well as reducing the risk of features being established on the untextured parts of the cloth. Finally, the contours of the biggest hole(s) are filled to define the texture mask.

During the third and final phase of initialization, we also address a prerequisite for the cloth recovery stage: the texture template. Firstly, the texture mask is applied to the original pre-processed frame, extracting the texture. The result is automatically cropped to the bounds of the content within the texture's mask, yielding the template. This process gives a known mapping for setting up point correspondences before tracking begins.

V. RECOVERY OF SPARSE 3D POINTS

This section focuses on reconstructing sparse 3D points from highly non-rigid 3D surfaces, such as cloth, captured in monocular video (or alternatively, a single image and a texture template). Figure 3 shows an overview of the approach. This work is in preparation for the next section which describes our contribution for a hierarchical reconstruction of a continuous surface from these sparsely recovered 3D surface points.

Tight clothing, such as *fitted T-shirts*, is characterized by locally near-rigid deformations which can be approximated by surface skinning techniques. We focus on the much more complex case of non-rigidly deforming cloth such as that represented by *standard T-shirts*. We assume that the cloth elasticity is negligible and that the texture on the T-shirt is large and screen-printed which gives locally near-inelastic properties, so that we can constrain the reconstruction algorithm to inelastic materials.

The first step involves establishing 2D point correspondences between the deforming surface in the current frame and the template. SIFT is often used for this purpose and is employed in [10], however it is not suitable for real-time. We employ simple real-time registration. Since we derive the

template from the image sequence itself, there is a known mapping between these images in world space. Robust feature points are established at regions of high curvature on the texture. These points are then tracked over the video sequence using sparse optical flow [19]. However, this optical flow algorithm traditionally finds pixel displacements in the sub-pixel range, and we require large displacements since we are capturing human movement from a webcam. Therefore, a multiresolution scheme on a Gaussian image pyramid is integrated to handle larger pixel displacements. Also, to handle occasional loss of features, we use temporal coherence to attempt to re-establish them. This simple method has limitations in that it does not regularize the optical flow field and does not support self-occlusion handling, so it is best suited to short videos.

For very fast recovery of sparse 3D points from the tracked 2D point correspondences we use the approach presented by Perriollat et al. [10]. The planar texture template is deformed to the unknown 3D surface by an unknown isometric transformation. Due to the inextensibility constraint, the geodesic curve between two points on the surface has the same length as the geodesic distance in the template, however, the surface deformation causes the Euclidean distance between the 3D points to be less. To find depths of the 2D points, their method computes bounds in a pairwise fashion where two points along with the inextensibility constraint bound the position of these points along their sightlines which intersect at the center of the projective camera. The overall set of upper bounds are iteratively refined to find an optimal solution.

VI. RECOVERY OF 3D CLOTH SURFACE

In this section, we propose a method to reconstruct a continuous 3D cloth surface from a limited set of recovered sparse 3D points.

A coarse surface mesh ξ_l for the local model is represented by transferring a regular mesh from the template with a mapping composed of three 2D to 1D thin-plate splines for distance to the depth bounds, the geometric smoothness, and the temporal smoothness.

By design, the surface recovered by the local model is most accurate at regions on the texture where many features are present. Geometric surface errors can be caused by interpolation across relatively untextured regions with few features and extrapolation outside of an arbitrary shaped texture. We propose to alleviate these errors by introducing a global model to infer some knowledge of the worn garment. As we focus on recovering a cloth surface and our demonstration is in respect to T-shirts, we choose to employ a cloth model for this purpose. Additionally, a cloth model gives us a *dynamic* surface recovery which can attempt to retain the real cloth behavior. This could have potential uses for novel future applications.

We design our global model around the mass-spring approach [20] since it is popular, efficient, and can achieve an accurate simulation of cloth [21]. Masses are arranged in a rectangular grid and connected to their neighbors by springs.

The vertical and horizontal springs constrain cloth stretching and compression whereas the diagonal springs constrain bending. We define simulation constraints for cloth stretching, bending, self collisions, edges, and deforming key vertices to the local recovery mesh. We present each constraint $C_{...}$ as a function with respect to a subset of vertex positions along with a weight parameter $k_{....}$. The cloth constraints are combined with their corresponding weights into one energy function which can be minimized with the Newton-Raphson method. Vertex positions are computed at least once every frame capture.

The procedure for initializing the global cloth mesh ξ_g is now described. We start by finding the approximate y coordinates of the neck and bottom of the T-shirt by minimizing and maximising y across a range of x coordinates at the center of the frame's cloth segmentation mask. Then we iterate upwards from the bottom of the shirt, minimizing and maximising x. We stop and define the bottom of each of the sleeve intersections once the change in x on each side of the shirt reaches a predefined threshold and continues to do so for at least 3 more y iterations, thus adding robustness to noise. The mesh edges are first fitted to the rectangle obtained from the x and y optimizations. The inner topology of ξ_g is structured so that it has a vertex aligned with each of the vertices in the initialized ξ_l, but at twice the resolution. The remaining inner vertices are established horizontally and vertically in proportion according to twice the ξ_l vertices/texture-width and vertices/texture-height ratios, respectively. Vertices on the mesh edges below the sleeve intersections are then refined to the edges of the segmentation mask. The x values for vertices on the left and right sides above the intersections are set to the x coordinate of each intersection respectively.

After initialization, the sleeve intersections continue to be computed every frame for defining the edge constraint C_{edge}.

Since we assume cloth elasticity to be negligible, we penalise compression and stretching of vertices which are horizontally or vertically adjacent. The stretching constraint is evaluated as:

$$C_{stretch}(\mathbf{v_i}, \mathbf{v_{i+1}}) = \| \mathbf{v_i} - \mathbf{v_{i+1}} \| - l_{i,i+1} \qquad (1)$$

where $l_{i,i+1}$ is the rest length of the edge between $\mathbf{v_i}$ and $\mathbf{v_{i+1}}$ in the initial state of the mesh. The bending constraint takes the form:

$$C_{bend}(\mathbf{v_i}, .., \mathbf{v_{i+3}}) = \arccos(\mathbf{n_l} \cdot \mathbf{n_r}) - \theta \qquad (2)$$

where $\mathbf{n_l}$ and $\mathbf{n_r}$ are the triangle normals of $\mathbf{v_i}$, $\mathbf{v_{i+1}}$, $\mathbf{v_{i+2}}$ and $\mathbf{v_{i+3}}$ and θ is the rest angle between these triangle normals. We use spatial hashing to handle self-collisions between the cloth according to:

$$C_{selfcol}(\mathbf{q}, \mathbf{v_i}, \mathbf{v_{i+1}}, \mathbf{v_{i+2}}) = (\mathbf{q} - \mathbf{v_i}) \cdot \mathbf{n_l} - h \qquad (3)$$

ensuring that point q remains above the triangle face $(\mathbf{v_i}, \mathbf{v_{i+1}}, \mathbf{v_{i+2}})$ with normal $\mathbf{n_l}$ by the cloth thickness h. The anchor constraint deforms the key anchor points in the global mesh to their corresponding points in the local mesh:

$$C_{anchor}(\mathbf{v_i}) = (\tilde{\mathbf{v}}_\mathbf{i} - \mathbf{v_i})^2, \forall i \in \Gamma \qquad (4)$$

where $\tilde{\mathbf{v}}_\mathbf{i}$ are the vertices in local mesh ξ_l and Γ is a set establishing which vertices in ξ_g should *anchor* to their corresponding points in ξ_l.

There is one external force comprising of gravity $\mathbf{f_g}$. There are three main parameters for the the global cloth model. The scalars k_{bend} and $k_{stretch}$ are defined empirically to attempt to mimic behavior of 100% cotton T-shirt cloth. The weight for C_{anchor} is a vector $\mathbf{k_{anchor}}$. Each weight $k_{anchor_i}, \forall i \in \Gamma$, is characterized by the proximity of point correspondences in the local neighborhood $\aleph_i$ of the corresponding vertex $\tilde{v}_i$ for the initial state of mesh ξ_l in the template where $k_{anchor_i} \propto \aleph_i$. We set k_{anchor_i} to zero for $i \notin \Gamma$.

Finally, patch tessellation is employed to subdivide the global surface ξ_g, increasing surface smoothness and realism. The rectangular region of ξ_g which corresponds to the local surface ξ_l is used for our reconstructions.

VII. Illumination Recovery and Rendering

The application of illumination recovery within our framework is two-fold. Rather than simply overlaying a new texture, we wish to remove the existing texture and then apply a new texture of a potentially different shape/size/color. We recovered the plain T-shirt color during segmentation initialization, so this leaves the illumination to be recovered in order to reconstruct the appearance of parts of the cloth which are covered by the real texture. Secondly, we modulate the alpha channel of the virtual texture with the recovered illumination to increase realism of the augmentation within the real scene. Given the segmentation mask of the printed texture which we have obtained, we can treat the mask as unknown pixel values and interpolate to reconstruct the illumination values, yielding recovery of the plain T-shirt with the printed design removed. Inpainting is an advanced interpolation scheme used to reconstruct unknown or lost parts of an image. We employ the fast inpainting approach presented by Telea [22] and downscale the input for speed. Figure 2 shows the illumination channels before and after illumination recovery by inpainting. This is a reasonable result which shows that high frequencies across the texture are smoothed out with negligible artifacts present.

OpenGL is employed to render the output according to the following algorithm:

1) The captured frame is rendered in the output window.
2) The virtual T-shirt color which the user selected in the GUI during initialization fills the T-shirt segmentation mask. Note that if the user does not wish to recolor the T-shirt, the color which was recovered earlier is used.
3) The virtual T-shirt image which the user selected in the GUI during initialization is mapped to the recovered 3D surface mesh.
4) The T-shirt segmentation mask is applied to the recovered illumination (i.e. the interpolated illumination channel for an untextured T-shirt) and used to modulate the alpha channel of the virtual T-shirt color and texture, hence increasing the realism of the augmentation within the real scene.
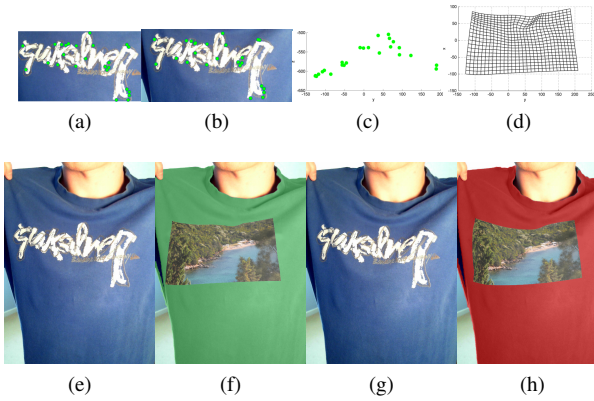
(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 4: Results for 3D Cloth Recovery and Retexturing.

VIII. Results and Discussion

Figure 2 shows good visual accuracy for our segmentation approach. Robustness is shown for arbitrary T-shirt colors and textures, subjects, and backgrounds.

The first row of Figure 4 shows intermediate reconstruction results consisting of tracking 2D point correspondences between a template derived from the start of the video sequence (initialization step) and an arbitrary frame with cloth deformation; sparse 3D point recovery; and continuous 3D surface recovery. The reconstructed surface mesh is convincing. Note that due to the way the camera has been calibrated, the more negative the points are along the *z-axis*, the closer they are to the camera.

The second row shows the results of recoloring and retexturing a T-shirt from two successive frame captures. The intermediate results are associated with the first example. Our augmented reality framework achieves photorealistic results at an average frame rate of 12fps on an Intel Core i7 CPU (4 cores at 2.93GHz).

Discussion: No specific or expensive hardware is required as we use a mid-range off-the-shelf PC and webcam under uncontrolled domestic lighting; whereas previous work appears to use a high-end camera along with special lighting to flatten the illumination. Our approach is subject to some limitations: (1) we assume that the face is visible in the input image (or in at least the first few frames of a video sequence), (2) there are no objects of a similar hue to the shirt in direct contact with it from a 2D frame perspective, and (3) the shirt is a plain-colored T-shirt containing a printed texture.

IX. Conclusion

We have presented a robust automated method for garment cloth and texture segmentation, and a hierarchical approach for 3D geometric reconstruction of highly deformable surfaces, such as cloth, which is robust to partially untextured regions given consecutive monocular video frames or a single image and a texture template. A real-time retexturing and recoloring framework has been demonstrated for combining these two methods for the purpose of augmented reality T-shirt try on.

Our results show visually accurate segmentation, convincing 3D shape reconstruction and photorealistic retexturing whilst employing a setup which is practical for a consumer.

For future work, we intend to extend our global surface reconstruction model and quantitatively analyse results.

References

[1] N. M. Thalmann, E. Lyard, M. Kasap, and P. Volino, "Adaptive Body, Motion and Cloth," in *Motion in Games*. Springer, Nov. 2008, p. 71.

[2] V. Gay-Bellile, A. Bartoli, and P. Sayd, "Deformable Surface Augmentation in Spite of Self-Occlusions," in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2007, pp. 1–4.

[3] J. Pilet, V. Lepetit, and P. Fua, "Fast non-rigid surface detection, registration and realistic augmentation," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 109–122, 2008.

[4] D. Bradley, G. Roth, and P. Bose, "Augmented reality on cloth with realistic illumination," *Machine Vision and Applications*, vol. 20, no. 2, pp. 85–92, 2009.

[5] A. Hilsmann and P. Eisert, "Tracking and Retexturing Cloth for Real-Time Virtual Clothing Applications," in *MIRAGE 2009*, 2009, p. 94.

[6] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of Multi-View stereo reconstruction algorithms," in *CVPR '06*, 2006, pp. 519–528.

[7] D. Bradley, T. Popa, A. Sheffer, W. Heidrich, and T. Boubekeur, "Markerless garment capture," in *ACM SIGGRAPH 2008 papers*. ACM, 2008, pp. 1–9.

[8] K. S. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popović, and S. M. Seitz, "Estimating cloth simulation parameters from video," in *ACM SIGGRAPH SCA '03*, 2003, pp. 37–51.

[9] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[10] M. Perriollat, R. Hartley, and A. Bartoli, "Monocular template-based reconstruction of inextensible surfaces," *International Journal of Computer Vision*, pp. 1–14, 2010.

[11] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres, "Monocular template-based reconstruction of smooth and inextensible surfaces," *Computer Vision–ACCV 2010*, pp. 52–66, 2011.

[12] M. Salzmann, F. Moreno-Noguer, V. Lepetit, and P. Fua, "Closed-form solution to non-rigid 3d surface registration," in *ECCV*, 2008, pp. 581–594.

[13] Y. Schnitman, Y. Caspi, D. Cohen-Or, and D. Lischinski, "Inducing semantic segmentation from an example," in *ACCV 2006*. Springer, 2006, pp. 373–384.

[14] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," in *CVPR 2008*. IEEE, 2008, pp. 1–8.

[15] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *Computer Graphics and Applications, IEEE*, vol. 21, no. 6, pp. 34–47, 2001.

[16] P. Viola and M. J. Jones, "Robust Real-Time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[17] R. D. Seely, "On a three-dimensional gait recognition system," Ph.D. dissertation, School of Electronics and Computer Science, University of Southampton, UK, 2010.

[18] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International joint conference on artificial intelligence*, vol. 3, 1981, pp. 674–679.

[20] M. Muller, B. Heidelberger, M. Hennix, and J. Ratcliff, "Position based dynamics," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 109–118, 2007.

[21] D. H. House and D. E. Breen, *Cloth modeling and animation*. AK Peters, Ltd. Natick, MA, USA, 2000.

[22] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004.