

Non-rigid Photometric Stereo with Colored Lights

Carlos Hernández¹ George Vogiatzis¹ Gabriel J. Brostow² Bjorn Stenger¹ Roberto Cipolla²
Computer Vision Group, Toshiba Research Europe, Cambridge, UK¹
Dept. of Engineering, University of Cambridge, Cambridge, UK²

Abstract

We present an algorithm and the associated capture methodology to acquire and track the detailed 3D shape, bends, and wrinkles of deforming surfaces. Moving 3D data has been difficult to obtain by methods that rely on known surface features, structured light, or silhouettes. Multispectral photometric stereo is an attractive alternative because it can recover a dense normal field from an un-textured surface. We show how to capture such data and register it over time to generate a single deforming surface.

Experiments were performed on video sequences of un-textured cloth, filmed under spatially separated red, green, and blue light sources. Our first finding is that using zero-depth-silhouettes as the initial boundary condition already produces rather smoothly varying per-frame reconstructions with high detail. Second, when these 3D reconstructions are augmented with 2D optical flow, one can register the first frame's reconstruction to every subsequent frame.

1. Introduction

The modeling of dynamic cloth geometry is now receiving considerable attention [18, 19, 21, 26, 27]. The complexity underlying the simplest of cloth motions motivates capturing cloth geometry and motion data from the real world.

Existing algorithms one might employ for capturing detailed 3D models of moving cloth include multiple view stereo [22], photometric stereo [9, 23], and laser based methods [12]. However, most of these techniques require that the subject stand still during the acquisition process, or move slowly [15]. Another substantial challenge is that even starting from a sequence of 3D scans of the deforming object, registration is necessary to produce a single deforming 3D model, suitable for CG animation or further data analysis.

The proposed technique for acquiring complex motion data from real moving cloth uses a practical setup that consists of an ordinary video camera and three colored light sources (see Fig. 1). The key observation is that in an en-

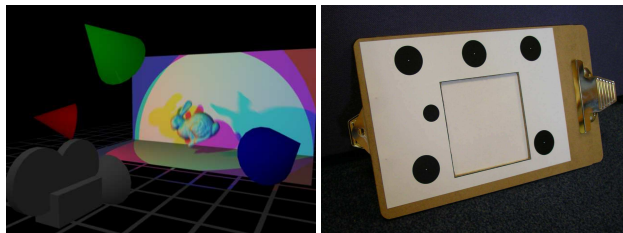


Figure 1. **Setup and calibration board.** Left: a schematic representation of our multispectral setup. Right: Linking the two clipboards and attaching a printed calibration pattern produces a planar trackable target for computing the orientation of the pattern's plane. Cloth inserted between the boards appears in the square hole and creates the association between color and orientation.

vironment where red, green, and blue light is emitted from different directions, a Lambertian surface will reflect each of those colors simultaneously without any mixing of the frequencies. The quantities of red, green and blue light reflected are a linear function of the surface normal direction. A color camera can measure these quantities, from which an estimate of the surface normal direction can be obtained. By applying this technique to a video sequence of a deforming object, one can obtain a sequence of normal maps for that object which are used in two distinct ways: (1) These normal maps are integrated to produce a sequence of depth-maps and (2) 2D optical flow is computed between consecutive frames of the normal map sequence. This optical flow, combined with a local rigidity constraint, registers the first depth map will all subsequent ones, producing a single deforming 3D model.

The main contributions of this work are the following:

1. A simple and practical acquisition setup for acquiring high-detail, per-frame reconstructions.
2. An optical-flow based registration technique for tracking the folds and creases of real moving cloth.
3. An interesting application of our method is a simple technique for 'dressing' a virtual character with real moving cloth.

The rest of the paper is laid out as follows: Section 2 discusses related prior work. In Section 3 we describe the extraction of a depth map sequence from video of a deforming object, while Section 4 presents the surface registration technique. We report on experiments we carried out to verify our approach in Section 5, and Section 6 concludes with a discussion of our main contributions.

2. Prior work

The animation and capture of cloth deformations is being explored in many fields, so we provide a general explanation of their relevance in the context of the proposed technique.

Texture Cues White and Forsyth [26, 27] and Scholz *et al.* [21] have presented work on using texture cues to perform the specific task of cloth capture. Their methods are based on printing a special pattern on a piece of cloth and capturing video sequences of that cloth in motion. The estimation of the cloth geometry is based on the observed deformations of the known pattern as well as texture cues extracted from the video sequence. The techniques produce results of very good quality but are ultimately limited by the requirement of printing a special pattern on the cloth which may not be practical for a variety of situations. In the present work, we avoid this requirement while producing detailed results.

Pilet *et al.* [18] and Salzmann *et al.* [19] proposed a slightly more flexible approach where one uses the pattern already printed in a piece of cloth, by presenting it to the system in a flattened state. Using sparse feature matching the pattern can be detected in each frame of a video sequence. Due to the fact that detection occurs separately in each frame, the method is quite robust to occlusions. However the presented results dealt only with minor non-rigid deformations.

Photometric Stereo Photometric stereo [28] is one of the most successful techniques for surface reconstruction from images. It works by observing how changing illumination alters the image intensity of points throughout the object surface. These changes reveal the local surface orientations. This field of local surface orientations can then be integrated into a 3D shape. State of the art photometric-stereo allows uncalibrated light estimation [14, 23] as well as multiple unknown albedos [7, 10]. The main difficulty with applying photometric stereo to deforming objects lies in the requirement of changing the light source direction for each captured frame, while the object remains still. This is quite impractical when reconstructing the 3D geometry of a *moving* object. We show how multispectral lighting allows one to essentially capture three images (each with a

different light direction) in a single snapshot, thus making per-frame photometric reconstruction possible.

Colored and Structured Lights The earliest related works are also the most relevant. The first reference to multispectral light for photometric stereo dates back 20 years to the work of Petrov [17]. Ten years later, Kontsevich *et al.* [11] actually demonstrated an algorithm for calibrating unknown color light sources and at the same time computing the surface normals of an object in the scene. They verified the theory on synthetic data and an image of a real egg. We use a simplified approach for calibration and the same orientation-from-color cue to eventually convert video of un-textured cloth into a single surface with complex changing deformations.

More recently, the parameters needed to simulate realistic cloth dynamics were measured in video by projecting explicitly structured horizontal light stripes onto material samples under static and dynamic conditions [3]. This system measured the edges and silhouette mismatches present in real vs. simulated sequences. Many researchers have utilized structured lighting, and Gu *et al.* [8] even used color, although their method is mostly for storing and manipulating acquired surface models of shading and geometry. Weise *et al.* [25] is the current state-of-the-art for structured light and has some advantages in terms of absolute 3D depth, but at the expense of both spatial and temporal sampling, e.g. 17 Hz compared to our 60 Hz (or faster, limited only by the camera used). Zhang *et al.* [29] is a nice complete system also with structured lighting that applies to face models and videos. Sand *et al.* dispensed with special lighting but leveraged motion capture and automatic silhouettes to deform a human body template [20]. Our technique, on the other hand, expects no prior models of the cloth being reconstructed.

3. Depth-map video

In this section, we follow the exposition of Kontsevich *et al.* [11]. For simplicity, we first focus on the case of a single distant light source with direction \mathbf{l} illuminating a Lambertian surface point P with surface orientation direction \mathbf{n} . Let $S(\lambda)$ be the energy distribution of that light-source as a function of wavelength λ and let $\rho(\lambda)$ be the spectral reflectance function representing the reflectance properties at that surface point. We assume our camera pixels consist of three sensors sensitive to different parts of the spectrum. If $\nu_i(\lambda)$ is the spectral sensitivity of the i -th sensor for the pixel that receives light from P , then intensity measured at that sensor is

$$r_i = \mathbf{l} \cdot \mathbf{n} \int S(\lambda) \rho(\lambda) \nu_i(\lambda) d\lambda, \quad (1)$$

or in matrix form

$$\mathbf{r} = M\mathbf{n}, \quad (2)$$

where the (i, j) th element of M is

$$m_{ij} = l_j \int S(\lambda)\rho(\lambda)v_i(\lambda)d\lambda. \quad (3)$$

When more light sources are added, if the system is linear and $\mathbf{l} \cdot \mathbf{n} \geq 0$ still holds for each light, the response of each sensor is just a sum of the responses for each light source individually, leading to equation (2) still being valid with

$$M = \sum_k M^k, \quad (4)$$

where M^k describes the k -th light source. Since each of the M^k is of rank 1, this implies that in the absence of self occlusions, a minimum of three different lights needs to be present in the scene for M to be invertible. If the surface is uniformly colored, then $\rho(\lambda)$ and consequently M will be fixed across all un-occluded locations.

Equation (2) establishes a 1-1 mapping between an RGB pixel measurement from a color camera and the surface orientation at the point projecting to that pixel. Our strategy is to use the inverse of this mapping to convert a video of a deformable surface into a sequence of normal maps. The next section describes our setup and explains how the RGB-to-normal mapping is estimated.

3.1. Setup and calibration

Our setup consists of a color video camera and three light sources which have been filtered with red, green and blue filters respectively. The camera is placed 5m away from the target object. The light sources are at a similar distance, aimed at the target and at an angle of about 30 degrees to one another, not colinear. The filming occurs in a dark room with minimal ambient light. Figure 1 (left) describes this schematically.

In [11] and [6], methods were proposed for the estimation of the linear mapping M of equation (2) from the image itself, using the constraints of uniform albedo and surface integrability that must be satisfied by the normal map \mathbf{n} . However the results obtained with these techniques can be unsatisfactory, especially in situations where the target object does not have many different surface orientations (if for example it is mostly planar). We prefer to estimate the mapping by employing an easy-to-use calibration tool (Figure 1 (right)) similar to the one used in [16]. The pattern is planar with special markings that allow the plane orientation to be estimated. By placing the cloth in the center of the pattern, we can measure the color it reflects at its current orientation. We thus obtain a sequence of (\mathbf{r}, \mathbf{n}) pairs to which we fit the mapping M using linear least squares.

3.2. Depth from Normals

By estimating and then inverting the linear mapping M linking RGB values to surface normals, we can convert a video sequence captured under colored light into a *video of normal-maps*. Due to the dark room conditions, by simple intensity thresholding we can segment background pixels in every frame of the original video, as they are almost perfectly black. We then integrate each normal map independently to obtain a depth map in every frame by imposing that the occluding contour is always at zero depth. This integration process is a fairly established technique and several algorithms are available. We have used the Successive Overrelaxation solver (SOR) [5] because of its robustness and simplicity. At the end of the integration process, we obtain a *video of depth-maps*.

4. Tracking the surface

While the *video of depth-maps* representation can be adequate for some applications, for cloth modeling, points on different depth maps must be put in correspondence. This is easily seen in the very common task of texture-mapping a moving cloth. Figure 3 (second row) shows the failure of directly texture-mapping each depth-map without any registration. Our approach is to use the first depth-map of the sequence as a 3D template which will be deformed to match all subsequent depth-maps. Let $z^k(u, v)$ denote the depth-map at frame t . Our deformable template, which corresponds to the depth-map at frame 0, is a triangular mesh with vertices

$$\mathbf{x}_i^0 = (u_i^0, v_i^0, z^0(u_i^0, v_i^0)) \quad i = 1 \dots N, \quad (5)$$

and a set of edges \mathcal{E} . At frame t the template mesh is deformed to fit the t -th depth-map by applying a translation \mathbf{T}_i^t to each vertex \mathbf{x}_i so the i -th vertex at frame t moves to $\mathbf{x}_i^0 + \mathbf{T}_i^t$. This approach follows [2] where a human body template is deformed to fit range scan data. The deformations of the template will be guided by the following two competing constraints:

- the deformations must be compatible with the frame-to-frame 2D optical flow of the original video sequence,
- the deformations must be locally as rigid as possible.

In the following sections we describe these two constraints in more detail.

4.1. 2D Optical flow

We begin by computing frame-to-frame optical flow in the video of normal-maps. A standard optical flow algorithm is used for this computation [4] which for every pixel

location (u, v) in frame t predicts the displacement $\mathbf{d}^t(u, v)$ of that pixel in frame $t + 1$. Let (u^t, v^t) denote the position in frame t of a pixel which in frame 0 was at (u^0, v^0) . We can *advect* $\mathbf{d}^t(u, v)$ to estimate (u^t, v^t) using the following equation from [20]

$$(u^j, v^j) = (u^{j-1}, v^{j-1}) + \mathbf{d}^{j-1}(u^{j-1}, v^{j-1}) \quad j = 1 \dots t. \quad (6)$$

If there was no error in the optical flow and our template from frame 0 had been perfectly deformed to match frame t , then vertex \mathbf{x}_i^0 of the template should be displaced to point

$$\mathbf{y}_i^t = (u_i^t, v_i^t, z^t(u_i^t, v_i^t)) \quad (7)$$

in frame t . We formulate the constraint as an energy term consisting of the sum of squared differences between the displaced vertex locations $\mathbf{x}_i^0 + \mathbf{T}_i^t$ and the positions predicted by the advected optical flow \mathbf{y}_i^t at frame t

$$E_D(\mathbf{T}_1^t, \dots, \mathbf{T}_N^t) = \sum_{i=1}^N \|\mathbf{x}_i^0 + \mathbf{T}_i^t - \mathbf{y}_i^t\|^2. \quad (8)$$

4.2. Rigidity

Simply moving each template vertex to the 3D position predicted by optical flow using (8) can cause stretching and other geometric artifacts like the ones displayed in Figure 3 (third row). This is due to error in the optical flow caused by image noise or occlusions. To regularize the deformation of the template mesh, we require translations applied to nearby vertices to be as similar as possible. This is achieved by an energy term E_R defined by

$$E_R(\mathbf{T}_1^t, \dots, \mathbf{T}_N^t) = \sum_{(i,j) \in \mathcal{E}} \|\mathbf{T}_i^t - \mathbf{T}_j^t\|^2. \quad (9)$$

4.3. Optimization

The two terms defined above are then combined into a single energy function

$$E_{TOT}(\mathbf{T}_1^t, \dots, \mathbf{T}_N^t) = \alpha E_D + (1 - \alpha) E_R, \quad (10)$$

which is optimized with respect to $\mathbf{T}_1^t, \dots, \mathbf{T}_N^t$ for every frame t . For the optimization we use an iterated scheme where we replace each \mathbf{T}_i^t with the optimal translation $\hat{\mathbf{T}}_i^t$ given that every other translation is constant. This is actually a quadratic minimization problem which has a closed-form solution leading to the update formula

$$\hat{\mathbf{T}}_i^t = \alpha (\mathbf{y}_i^t - \mathbf{x}_i) + (1 - \alpha) \frac{1}{\mathcal{N}(i)} \sum_{j \in \mathcal{N}(i)} \mathbf{T}_j^t, \quad (11)$$

where $\mathcal{N}(i)$ is the set of neighbors of vertex i and α is a parameter determining the degree of rigidity of the mesh. The next section describes a set of experiments we carried out to verify the quality of reconstruction and tracking of deforming cloth.

5. Experiments

In all the results presented here we have used a color video camera with a resolution of 1280×720 and 60 fps. Computation times are in the order of 20 seconds per frame for the depth-map recovery and another 20 seconds for the registration of the template mesh to the current frame. All computations were carried out on a 2.8Ghz Pentium 4 processor with 2Gb of RAM.

5.1. Comparison with photometric stereo

To evaluate the accuracy of the per-frame depth-map estimation we reconstructed a static object (a jacket) using classic photometric stereo with three images each taken under different illumination. The same object was reconstructed using a single image, captured under simultaneous illumination by three colored lights, using our technique. Figure 2 shows the two reconstructions side by side. The results look very similar and the average distance between the two meshes is only **1.4%** of the bounding box diagonal. This demonstrates that equation (2) works well in practice. It is worth noting that even though photometric stereo achieves comparable accuracy, it cannot be used on a non-static object whose shape will change while the three different images are captured. Since our method only uses one image, it is suitable for obtaining frame-by-frame reconstructions of a deforming object.

5.2. Texture-mapped moving cloth

For the second experiment shown here, a model wearing a white sweater was filmed dancing under our multispectral illumination setup (see first row of Figure 3). We used this sequence to evaluate the mesh registration algorithm of Section 4 by texture mapping the deforming sweater. Figure 3 shows several approaches to mesh registration starting with no registration at all (second row), registration using the advected optical flow alone (third row) and the effect of regularizing optical flow with the rigidity constraint (fourth row), which is what this paper proposes. This last approach is seen to outperform all others as it manages to track the surface for more than 500 frames.

In Figure 4 we show several views of frame 380 without the texture map in high resolution (the mesh consists of approximately 180,000 vertices). The images clearly show the high frequency detail of the sweater. To the best of our knowledge, this is the only method able to reconstruct deforming cloth with such detail.

5.3. ‘Dressing’ a virtual character with moving cloth

To demonstrate the potential of our method for capturing cloth for animation, we attach a captured moving mesh to an articulated skeleton. Skinning algorithms have varying

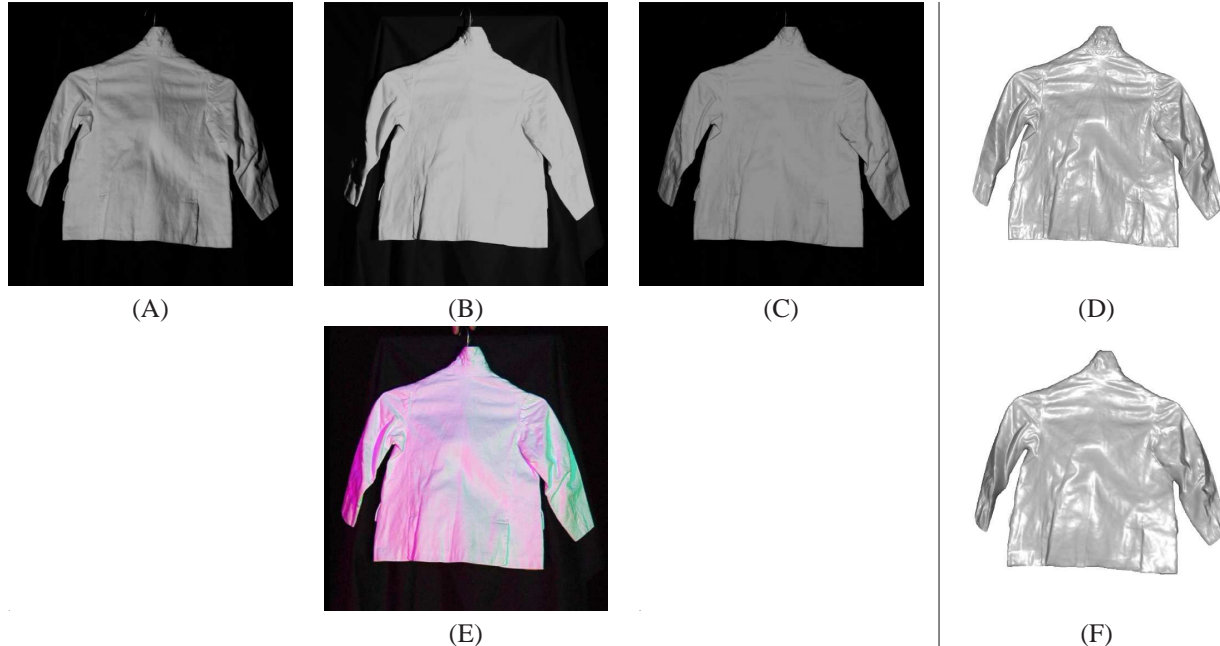


Figure 2. **Comparison with photometric stereo.** (A-C) show three grayscale images captured by a digital camera, each taken under a different illumination, providing the input to a classic photometric stereo reconstruction [28] shown in (D). (E) shows a frame from a jacket sequence, where the same object is illuminated *simultaneously* by three different colored lights. Our algorithm only uses one such frame to generate the surface mesh shown in (F). Note that both algorithms give very similar results, but only the new one (bottom row) can work with video since only one frame is required to obtain a reconstruction. As a quantitative comparison, the average error between both reconstructions is only **1.4%** of the bounding box diagonal.

degrees of realism and complexity, e.g. [13]. We apply a version of smooth skinning in which each vertex \mathbf{v}_k in the mesh is attached to one or more skeleton joints and a link to joint i is weighted by $w_{i,k}$. The weights control how much each joint i affects the transformation of the vertex [24]

$$\mathbf{v}_k^t = \sum_i w_{i,k} \mathbf{S}_i^{t-1} \mathbf{v}_k^{t-1}, \quad \sum_i w_{i,k} = 1, \quad (12)$$

where the matrix \mathbf{S}_i^t represents the transformation from joint i 's local space to world space at time instant t . The mesh is attached to the skeleton by first aligning both in a fixed pose and then finding, for each mesh vertex, a set of nearest neighbors on the skeleton. The weights are set inversely proportional to these distances. The skeleton is animated using publicly available mocap data [1] while the mesh is animated by playing back one of our captured and registered cloth sequences. Figure 5 shows example frames from the rendered sequence (please also see the video). Even though the skeleton and cloth motions are not explicitly aligned, the visual effect of the cloth moving on a controllable character is appealing. Such data-driven cloth animation can serve as a useful tool and presents an alternative to physical cloth simulation.

6. Conclusion

Building on the long established but surprisingly overlooked theory of multispectral lighting for photometric stereo, we have discovered and overcome several new obstacles. We developed a capture methodology that parallels existing work for capturing static cloth, but also enables one to capture the changing shape of cloth in motion. Integration of the resulting normal fields is already possible with the simple boundary condition that the occluding contour is at zero depth. We have verified the accuracy of the depth-maps against classic photometric stereo. When such a sequence of surfaces is played back, it appears to be changing smoothly. However, to really make use of such data, we had to explore long-term registration.

The high level of detail captured by the normal fields includes surface bends, wrinkles, and even temporary folds. Tracking of folds is inherently underconstrained, and will continue to be a challenge. The solution we propose is to register at least the visible details of the surface so that points on the surface in the first frame are tracked through to the last, through locally rigid deformations. In our framework, 2D optical flow is combined with the sequence of depth maps, creating a powerful 3D constraint which guides the tracking through noisy flow as well as severe deforma-

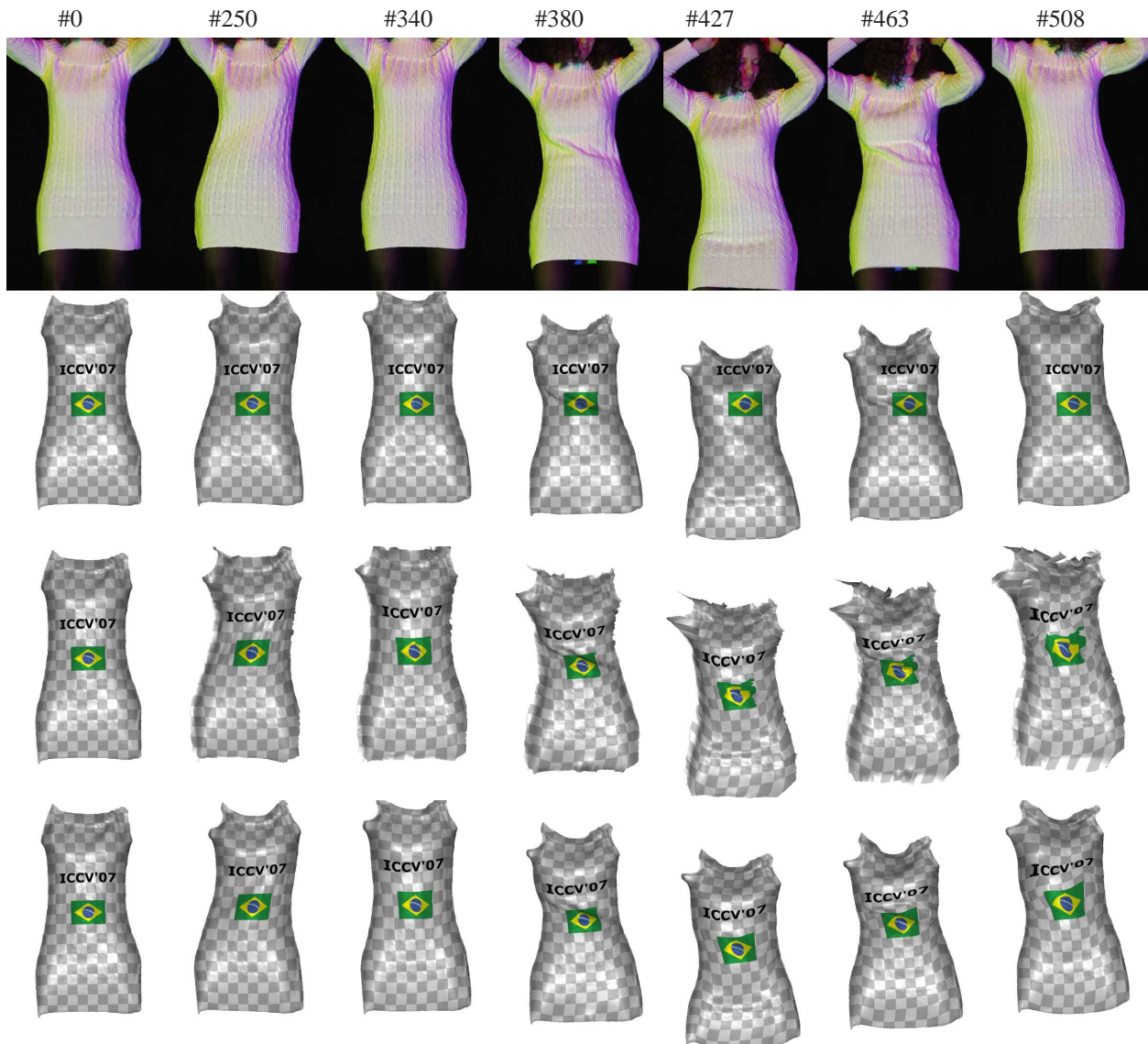


Figure 3. **Cloth tracking results of a sweater sequence.** First row: input video sequence of a person wearing a white sweater while being illuminated by three colored lights from three different orientations. Second row: *video of depth-maps* obtained by the technique described in Section 3 and directly texture mapped without any registration. The approach is quickly seen to fail after a few frames. Third row: texture-mapping is obtained by advecting frame-by-frame 2D optical flow [20]. Error in the optical flow advection causes artifacts after about 380 frames. Last row: Proposed method (Section 4) where 2D optical flow is regularized with a rigidity constraint to reduce advection errors. The texture is correctly registered throughout the entire sequence (please see the video).

tions. Our experiments, documented here and also in the supplementary video, demonstrate this.

We found it practical to trade the printing of intricate patterns on cloth for a slightly more complex lighting setup. We hope this will lower the barrier to entry for others wishing to record detailed cloth deformation. Finally, with access to a unique stream of rich 3D cloth poses, we have shown how easily the data is employed in a creative context for realistic character animation of a clothed avatar.

References

- [1] CMU Graphics Lab Motion Capture Database. <http://mocap.cs.cmu.edu>. 5
- [2] B. Allen, B. Curless, and Z. Popović. Articulated body deformation from range scan data. In *Proc. of the ACM SIGGRAPH*, pages 612–619, 2002. 3
- [3] 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 *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Com-*



Figure 4. **Cloth reconstruction results of a deforming sweater.** Multispectral photometric reconstruction of a single frame of a longer video sequence using the technique described in section 3. From left to right, multiple viewing angles (frontal, ± 25 degrees, ± 50 degrees) of frame 380 of the sweater sequence (see supplementary video for more details).

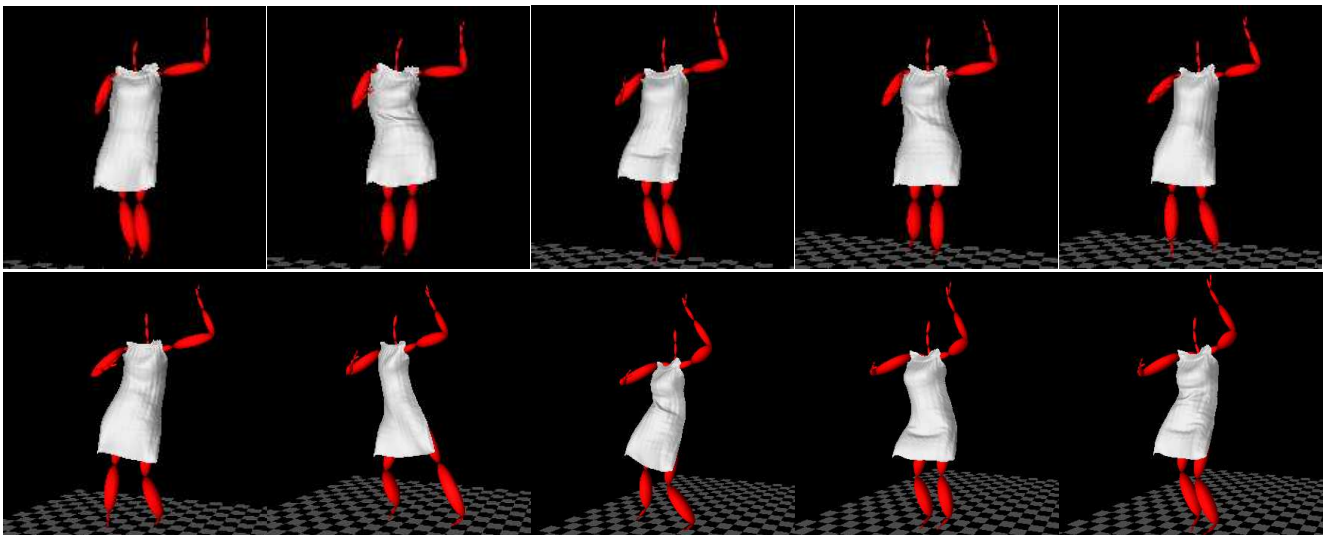


Figure 5. **Attaching captured moving cloth to an animated character.** We apply smooth skinning to attach a moving mesh to an articulated skeleton that can be animated with mocap data. The mesh is simply animated by playing back one of our captured and registered cloth sequences, in this case a dancing sequence (please also see the video).

- puter animation, pages 37–51, 2003. 2
- [4] M. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. In *Computer Vision and Image Understanding*, volume 63(1), pages 75–104, January 1996. 3
- [5] M. Davis and J. McCammon. Solving the finite difference linearized poisson-boltzmann equation: A comparison of relaxation and conjugate gradient methods. *J. Comput. Chem.*, 10(3):386–391, 1989. 3
- [6] M. S. Drew. Direct solution of orientation-from-color problem using a modification of Pentland’s light source direction estimator. *Comput. Vis. Image Underst.*, 64(2):286–299, 1996. 3
- [7] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying BRDFs from photometric stereo. In *ICCV ’05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05)*, volume 1, pages 341–348, 2005. 2
- [8] X. Gu, S. Zhang, P. Huang, L. Zhang, S.-T. Yau, and R. Martin. Holoimages. In *SPM ’06: Proceedings of the 2006 ACM Symposium on Solid and Physical Modeling*, pages 129–138. ACM Press, 2006. 2
- [9] A. Hertzmann and S. Seitz. Shape and materials by example: a photometric stereo approach. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages I: 533–540, 2003. 1
- [10] A. Hertzmann and S. Seitz. Shape reconstruction with general, varying BRDFs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1254–1264, November 2005. 2
- [11] L. Kontsevich, A. Petrov, and I. Vergelskaya. Reconstruction of shape from shading in color images. *J. Opt. Soc. Am. A*, 11(3):1047–1052, 1994. 2, 3
- [12] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The Digital Michelangelo Project: 3d

- scanning of large statues. In *Proc. of the ACM SIGGRAPH*, page 1522, 2000. [1](#)
- [13] J. P. Lewis, M. Cordner, and N. Fong. Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 165–172, 2000. [5](#)
- [14] J. Lim, J. Ho, M.-H. Yang, and D. Kriegman. Passive photometric stereo from motion. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1635–1642, 2005. [2](#)
- [15] T. Malzbender, D. G. B. Wilburn, and B. Ambrisco. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Eurographics Symposium on Rendering 2006*, Nicosia, Cyprus, June 2006. [1](#)
- [16] J. A. Paterson, D. Claus, and A. W. Fitzgibbon. BRDF and geometry capture from extended inhomogeneous samples using flash photography. *Computer Graphics Forum (Special Eurographics Issue)*, 24(3):383–391, Sept. 2005. [3](#)
- [17] A. Petrov. Light, color and shape. *Cognitive Processes and their Simulation (in Russian)*, pages 350–358, 1987. [2](#)
- [18] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. In *Conference on Computer Vision and Pattern Recognition, San Diego, CA*, June 2005. [1](#), [2](#)
- [19] M. Salzmann, S. Ilic, and P. Fua. Physically valid shape parameterization for monocular 3-d deformable surface tracking. In *British Machine Vision Conference*, 2005. [1](#), [2](#)
- [20] P. Sand, L. McMillan, and J. Popović. Continuous capture of skin deformation. *ACM Trans. Graph.*, 22(3):578–586, 2003. [2](#), [4](#), [6](#)
- [21] V. Scholz, T. Stich, M. Keckeisen, M. Wacker, and M. Magnor. Garment motion capture using color-coded patterns. *Computer Graphics Forum (Proc. Eurographics EG'05)*, 24(3):439–448, Aug. 2005. [1](#), [2](#)
- [22] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 519–528, 2006. [1](#)
- [23] G. Vogiatzis, C. Hernandez, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1847–1854, 2006. [1](#), [2](#)
- [24] X. C. Wang and C. Phillips. Multi-weight enveloping: least-squares approximation techniques for skin animation. In *SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 129–138, 2002. [5](#)
- [25] T. Weise, B. Leibe, and L. V. Gool. Fast 3d scanning with automatic motion compensation. In *CVPR '07: Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007. [2](#)
- [26] R. White and D. Forsyth. Combining cues: Shape from shading and texture. In *Computer Vision and Pattern Recognition*, volume 2, pages 1809–1816, 2006. [1](#), [2](#)
- [27] R. White and D. Forsyth. Retexturing single views using texture and shading. In *European Conference on Computer Vision*, volume LNCS 3954, pages 70–81. Springer, 2006. [1](#), [2](#)
- [28] R. Woodham. Photometric method for determining surface orientation from multiple images. In *Optical Eng.*, number 1, pages 139–144, 1980. [2](#), [5](#)
- [29] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz. Spacetime faces: High-resolution capture for modeling and animation. In *ACM Annual Conference on Computer Graphics*, pages 548–558, August 2004. [2](#)