# Cloth Capture

Ryan White Anthony Lobay D.A. Forsyth
UC Berkeley

**Overview** We present a method for capturing the geometry and parameterization of fast-moving cloth using multiple video cameras, without requiring camera calibration. Our cloth is printed with a multi-scale pattern that allows capture at both high speed and high spatial resolution even though self-occlusion might block any individual camera from seeing the majority of the cloth. We show how to incorporate knowledge of this pattern into conventional structure-from-motion approaches, and use a novel scheme for camera calibration using the pattern. We use bundle adjustment to obtain accurate reconstruction results and use the parameterization to drive several stages of post processing including strain reduction and a cloth simulation to fill in gaps. We demonstrate our algorithm by capturing, retexturing and displaying several sequences of fast moving cloth.

**Pattern Selection** Our cloth is printed with a pattern that contains information at multiple scales. At the larger scale, our pattern includes unique multi-colored triangles with parametric information about the location of the triangle in the domain of the cloth, as well as surface normals. Within each large triangle, there are 15 vertices that allow us to produce a fine mesh with a high degree of spatial accuracy. The pattern consists of areas of constant color to avoid problems with motion blur, a small number of unique colors to reduce confusion, and a pattern that can scale to hundreds of unique large triangles containing thousands of unique points.

**Localization** In each view, we perform a hierarchical search for points. First, we identify the locations and normals of the large triangles. The normals are resolved up to a two-fold ambiguity. Next, we estimate the colors of the triangle to determine where in the cloth it appears. Finally, we search for the locations of the corners of the smaller triangles.

**Camera Calibration** Extending work in the shape from texture literature [2004], we calibrate our cameras using both points and normals. Choosing two large triangles at random, we estimate an orthographic camera by using information from the locations of the triangles and the associated normals. We prove that, aside from an ambiguity in the direction of the normals, a metric reconstruction is available from only two points and two normals. We search over normal ambiguities and a number of randomly chosen triangle pairs to select the orthographic calibration that produces the lowest reconstruction error for the remaining triangles.

**3D Reconstruction** With a fairly accurate estimate of an orthographic camera model, we can reconstruct the locations of the triangles in each camera. However, when our cameras observe the scene from radically different angles, there are perspective effects. To compensate, we upgrade to a perspective camera model. We initialize assuming rough values from physical measurements and use nonlinear optimization (bundle adjustment) to compute more accurate camera parameters. Bundle adjustment reduces reprojection errors from above 10 pixels to less than 3 pixels.

**Post Processing** After reconstructing the 3D location and parametric value of each observed point on the cloth, there are minor errors: image noise in localization, temporal noise, errors in camera calibration and missing observations. While some of these errors can be corrected with simple smoothing and minor validation, regions of missing detection and disagreement between cameras present more significant problems. We use a cloth model to fill in gaps. Because gaps are small and well constrained, our cloth simulation does not suffer from typical stiffness and stability problems. We also use surface parameterization to drive a strain reduction step.

**Results** We capture and retexture several sequences that would be difficult to simulate due to complex environmental interactions, including mechanical contact and wind (shown in the figure). We filmed using four video cameras capturing at 30 Hz.

## References

LOBAY, A., AND FORSYTH, D. 2004. Recovering shape and irradiance maps from rich dense texton fields. In *Proceedings of Computer Vision and Pattern Recognition (CVPR)*.