Relightable 3D Video

Christian Theobalt[1], **Naveed Ahmed**[1], **Edilson de Aguiar**[1], **Gernot Ziegler**[1], **Hendrik Lensch**[2], **Marcus Magnor**[1], **Hans-Peter Seidel**[1]

[1]**MPI Informatik** [2]**Stanford University**

Overview

In order to import 3D videos of human actors into virtual environments, the appearance of the virtual human must be realistically rendered under the novel virtual lighting conditions. To achieve this, surface reflectance properties must be known. We have developed a model-based approach that captures human shape, motion and dynamic reflectance characteristics from a handful of synchronized video recordings.

Acquisition

- 8 synchronized video cameras (1004x1004 pixels; 25 fps; 12 bits per pixel).
- Camera arrangement: semi-circular setup around center; cameras are geometrically and photometrically calibrated.
- For each person and each type of apparel we record:
  - 1 Reflectance Estimation Sequence (RES): recorded in single-shot mode; person strikes initialization pose and rotates by approx. 5° around vertical body axis in-between subsequent frames; scene illuminated by one light source.
  - Several Dynamic Scene Sequences (DSS): person moves arbitrarily.

Reconstructing Shape and Motion

Silhouette-based analysis-by-synthesis [1]:

- Adaptable body model (16 segments, kinematic skeleton).
- Initialization: model's shape and proportions customized.
- Pose estimation for all frames in RES and DSS: silhouette-based motion capture.

Image-based Warp-Correction

Video frames and reflectance model parameters are stored in the texture domain. Each input video frame is transformed into a multi-view video (MVV) texture. During MVV texture generation we resample and align the input streams using a novel warp-correction scheme (Fig. 1) in order to compensate registration errors.

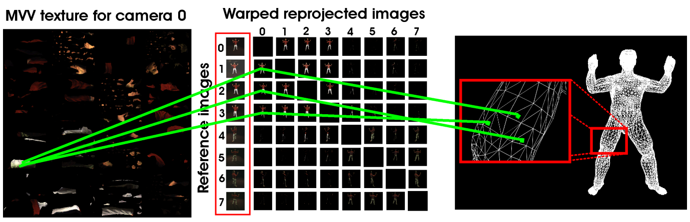

Figure 1: MVV texture generation for camera 0: The color information for each surface point on the body model is not looked up in the original input video frame recorded from camera 0. Instead the texel color is taken from an image that has been obtained by reprojecting the model that has been textured with camera image 0 into the camera view that sees the surface point most head-on.

Dynamic Reflectometry

BRDF Estimation

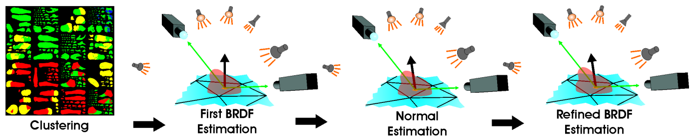

Figure 2: Subsequent steps to estimate per-texel BRDF.

Using the RES, an iterative optimization procedure estimates per-texel parametric BRDF models (Phong or Lafortune), Fig. 2. Our approach is an extension of [2] to time-varying data. For each texel, the employed energy function measures the sum of squared differences between the measured reflectance samples and their predicted appearances. Since the person rotates a large number of ingoing and outgoing light directions is considered.

Time-varying Normal Map Estimation

Given the BRDF models, a time-varying surface normal field is estimated from each DSS by means of photometric stereo (Fig. 3a).

Results

From each DSS, a relightable free-viewpoint video is reconstructed that can be rendered in real-time on standard graphics hardware (Fig. 3b,c).

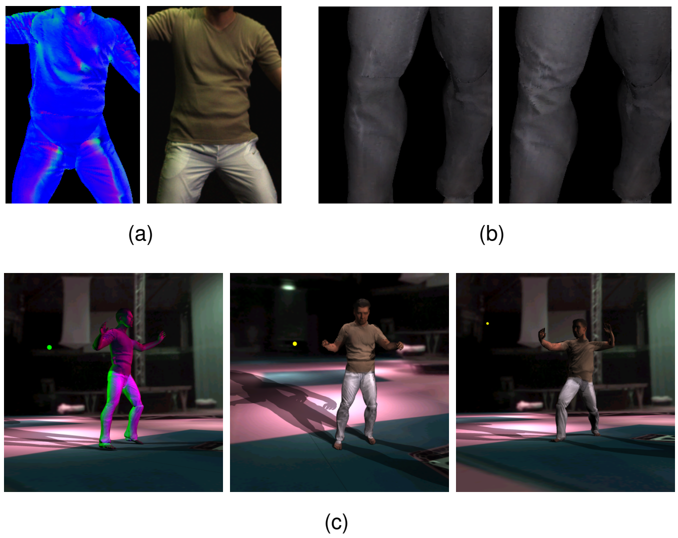

Figure 3: (a) Color-coded (local) normal map (l) and corresponding input video frame (r). (b) Rendered dynamic wrinkles in trousers. (c) Person rendered from different viewpoints and under different illumination conditions.

[1] J. Carranza, C. Theobalt., M. Magnor, H.P. Seidel, “Free-viewpoint Video of Human Actors”, Proc. SIGGRAPH ’03, p.569-577, 2003

[2] H. Lensch, J. Kautz, M. Goesele, W. Heidrich, H.P. Seidel: “Image-based reconstruction of spatial appearance and geometric detail”, ACM TOG, 22(2), p.234-257, 2003