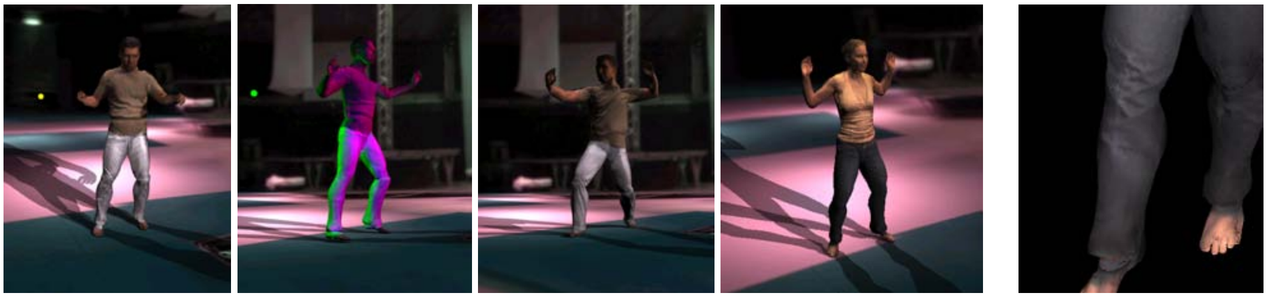
Joint Motion and Reflectance Capture for Relightable 3D Video

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, Germany [2]Stanford University, CA
{theobalt,nahmed,edeaguia,gziegler,magnor,hpseidel}@mpi-sb.mpg.de lensch@stanford.edu



(a) (b)

Figure 1: (a) 3D videos rendered under different novel lighting conditions. (b) Wrinkles in the trousers are captured in the normal field.

Keywords: 3D video, relighting, dynamic reflectometry

1 Introduction

3D Videos of Human Actors can be faithfully reconstructed from multiple synchronized video streams by means of a model-based analysis-by-synthesis approach [Carranza et al. 2003]. The reconstructed videos play back in real-time and the virtual viewpoint onto the scene can be arbitrarily changed. By this means, authentically animated, photo-realistically and view-dependently textured models of real people can be created that look real under fixed illumination conditions. To import real-world characters into virtual environments, however, relighting is vital and thus also surface reflectance properties must be known. To serve this purpose we have developed a video-based modeling approach that captures human motion as well as reflectance characteristics from a handful of synchronized video recordings. The presented method is able to recover spatially varying reflectance properties of skin and clothes by exploiting the time-varying orientation of each surface point with respect to camera and light direction. The resulting model description enables us to match animated subject appearance to different lighting conditions, as well as to interchange surface attributes among different people, e.g. for virtual dressing. Our relightable 3D videos allow populating virtual worlds with convincingly relit real-world people.

2 Exposition

The input to our system consists of multi-view video sequences that have been recorded with eight synchronized color video cameras, each of which features 1004x1004 pixel frame resolution, 12 bit per-pixel color depth, and a frame rate of 25 fps. For each person and each type of apparel two types of sequences are recorded. In the reflectance estimation sequence, RES, the set is illuminated with one spot light and the person rotates in place while attaining an initialization pose. From the RES surface reflectance properties will be estimated. The video footage, from which the actual relightable 3D video is reconstructed, is captured in a dynamic scene sequence, DSS, which can show any arbitrary body motion. A generic body model is adapted to match the shape and proportions of the recorded person. Subsequently, pose parameters are computed for all time frames of both types of video footage by means of a silhouette-based marker-free motion capture approach [Carranza et al. 2003]. To store all per-surface element data needed during reflectance estimation in texture space, we make use of a texture atlas as surface parameterization of the body model. Multi-view video (MVV) textures are generated by transforming each input video image into the texture domain. To correct for photo-inconsistencies due to approximate body geometry, we have developed an image-based warp-correction method that can be applied prior to MVV texture generation. Our dynamic reflectance model consists of two components: The first component is a parametric BRDF model for each point on the body surface. The second component is a time-varying normal map that incorporates dynamic changes in surface geometry, such as wrinkles in clothing. From the RES video data, BRDF model parameter values are estimated for each surface element (texel) of the geometry model individually. We optimize for optimal per-texel BRDF parameters by minimizing an energy functional that assesses the deviation between rendered and recorded appearance of the actor [Lensch et al.]. In our experiments we have employed the Phong model [Phong 1975] and the Lafortune model [Lafortune et al. 1997], but our approach is general enough to incorporate any parametric BRDF model. The recovered local BRDF parameters then allow us to estimate the time-varying surface normal field in the DSS sequences using the same energy minimization principle.

The animated body model, its spatially-varying reflectance, and the time-varying normal field enable us to interactively render and instantaneously relight the DSS sequences from arbitrary viewpoint and illumination direction on standard graphics hardware. Fig. 1 shows some results that we obtained with our approach.

References

CARRANZA, J., THEOBALT, C., MAGNOR, M., AND SEIDEL, H.-P. 2003. Free-viewpoint video of human actors. In *Proc. of SIGGRAPH'03*, 569–577.

LAFORTUNE, E., FOO, S., TORRANCE, K., AND GREENBERG, D. 1997. Non-Linear Approximation of Reflectance Functions. In *SIGGRAPH*, 117–126.

LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*.

PHONG, B.-T. 1975. Illumnation for computer generated pictures. *Communications of the ACM*, 311–317.