Spatio-temporal Reflectance Sharing for Relightable 3D Video

Naveed Ahmed, Christian Theobalt, and Hans-Peter Seidel

MPI Informatik, Saarbrücken, Germany,
[nahmed,theobalt,hpseidel]@mpi-inf.mpg.de,
WWW home page: http://www.mpi-inf.mpg.de/

Abstract. In our previous work [21], we have shown that by means of a model-based approach, relightable free-viewpoint videos of human actors can be reconstructed from only a handful of multi-view video streams recorded under calibrated illumination. To achieve this purpose, we employ a marker-free motion capture approach to measure dynamic human scene geometry. Reflectance samples for each surface point are captured by exploiting the fact that, due to the person's motion, each surface location is, over time, exposed to the acquisition sensors under varying orientations. Although this is the first setup of its kind to measure surface reflectance from footage of arbitrary human performances, our approach may lead to a biased sampling of surface reflectance since each surface point is only seen under a limited number of half-vector directions. We thus propose in this paper a novel algorithm that reduces the bias in BRDF estimates of a single surface point by cleverly taking into account reflectance samples from other surface locations made of similar material. We demonstrate the improvements achieved with this spatio-temporal reflectance sharing approach both visually and quantitatively.

1 Introduction

The capturing of relightable dynamic scene descriptions of real-world events requires the proper solution to many different inverse problems. First, the dynamic shape and motion of the objects in the scene have to be captured from multi-view video. Second, the dynamic reflectance properties of the visible surfaces need to be estimated. Due to the inherent computational complexity, it has not been possible yet to solve all these problems for general scenes. However, in previous work [21] we have demonstrated that the commitment to an adaptable a priori shape model enables us to reconstruct relightable 3D videos of one specific type of scene, namely of human actors. By means of a marker-free optical motion capture algorithm, it becomes possible to measure both the shape and the motion of a person from multiple synchronized video streams [2]. If the video footage has, in addition, been captured under calibrated lighting conditions, the video frames showing the moving person not only represent texture samples, but actually reflectance samples. Since a description of time-varying scene geometry

is at our disposition, we know under what different incoming light and outgoing viewing directions each point on the body surface is seen while the person is moving. It thus becomes feasible to fit to each point on the body surface a static parametric BRDF to describe the material properties, and a time-varying normal to describe the dynamic change in surface geometry. Although we have shown in our previous work that it is feasible to reconstruct dynamic surface reflectance properties using only eight cameras and a static set of light sources, this type of sensor arrangement leads to a biased sampling of the reflectance space. Due to the fixed relative arrangement of lights and cameras, each surface point is only seen under a limited number of half-vector directions. Furthermore, even if an actor performs very expressive motion, surface points will never be seen under all possible relative orientations to the cameras. This bias in the reflectance data leads to a bias in the measured reflectance models which may lead to an unnatural appearance if the 3D videos are rendered under virtual lighting conditions that are starkly different from the measurement setup. We thus propose spatio-temporal reflectance sharing, a method to cleverly combine dynamic reflectance samples from different surface points of similar material during BRDF estimation of one specific surface point. The guiding idea behind the approach is to exploit spatial coherence on the surface to obtain more samples for each texel while not compromising the estimation of spatially-varying details in surface appearance. Temporal coherence is also exploited, since samples are collected by combining measurements from subsequent time steps.

We continue with a review of the most relevant related work in Sect. 2. The acquisition setup, the model-based motion estimation approach, as well as the basic principles of dynamic reflectance estimation are described in Sect. 3. In Sect. 4, we describe the nuts and bolts of our proposed dynamic reflectance sharing approach and show how it fits into the original pipeline. Finally, in Sect. 5 we demonstrate both visually and quantitatively that our novel sampling strategy leads to improved results. We conclude in Sect. 6 and give an outlook to possible future work.

2 Related Work

There is a huge body of literature on the estimation of reflectance properties of static scenes from images that are captured under calibrated setups of light source and camera. Typically, parameters of a BRDF model are fitted to the data [18, 11] or appearance under novel lighting conditions is created via interpolation between the images themselves [15]. A combination of reflectance estimation and shape-from-shading to refine the geometry of static scenes is also feasible [25, 17, 1, 4, 5]. In an independent line of research, many methods to capture and render 3D videos of real-world scenes have been developed in recent years. A popular category of algorithms employs the shape-from-silhouette principle to reconstruct dynamic scene geometry by means of voxels, polyhedrals or point primitives [14, 24, 13, 6, 12]. By finding temporal correspondences between per-time-step reconstructions, it becomes feasible to generate novel animations as

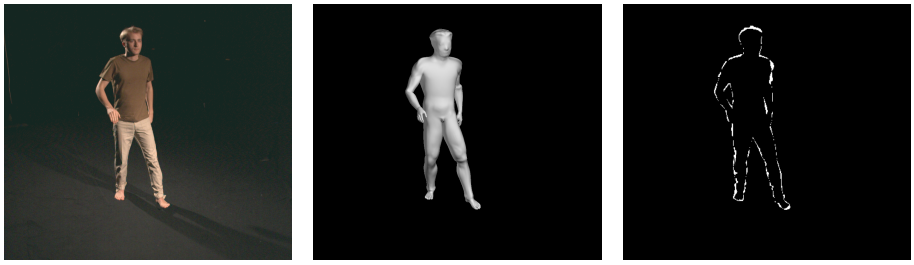
well [19]. Another category of approaches reconstructs dynamic scene geometry by means of multi-view stereo [27, 9, 22]. In any case, time-varying textures for rendering are assembled from the input video streams. In contrast, the authors in their previous work have proposed a model-based approach to free-viewpoint video of human actors that jointly employs a marker-free motion capture method and a dynamic multi-view texture generation approach to produce novel viewpoint renditions [2, 20]. Unfortunately, none of the aforementioned methods can correctly reproduce 3D video appearance under novel simulated lighting conditions.

Only few papers have been published so far that aim at relighting of dynamic scenes. In [8], a method to generate animatable and relightable face models from images taken with a special light stage is described. Wenger et al. [23] extend the light stage device such that it enables capturing of dynamic reflectance fields. Their results are impressive, however it is not possible to change the viewpoint in the scene. Einarsson et. al. [3] extends it further by using a large light stage, a trade-mill where the person walks on, and light field rendering for display. Eventually, human performances can be rendered from novel perspectives and relit. Unfortunately the method can only present single periodic motion, such as walking, and is only suitable for low frequency relighting.

In contrast, the authors have proposed a model-based method for reconstructing relightable free-viewpoint videos that extends measurement principles for static parametric reflectance models to dynamic scenes [21]. For our 3D video scenario, we prefer a compact scene description based on parametric BRDFs that can be reconstructed in a fairly simple acquisition facility. This paper proposes a novel solution to one important subproblem in the overall process, namely the clever sampling of surface reflectance in order to minimize the bias in the estimated BRDFs. This work has been inspired by the reflectance sharing method of Zickler et al. to reconstruct appearance of static scenes [26]. By regarding reflectance estimation as a scattered interpolation problem, they can exploit spatial coherence to obtain more reliable surface estimate. Our algorithm exploits both spatial and temporal coherence to reliably estimate dynamic reflectance. However, since a full-blown scattered data interpolation would be illusive with our huge sets of samples, we propose a faster heuristic approach to reflectance sharing.

3 Relightable Free-viewpoint Video of Human Actors - Preliminaries

The algorithm presented in this paper is a methodical improvement of one important step within a larger framework to reconstruct and render relightable free-viewpoint videos of human actors [21]. Although the algorithmic details of this framework, as a whole, are not the subject of this paper, for better understanding in the following we briefly elaborate on the acquisition setup used, as well as the employed model-based marker-less motion capture algorithm. Thereafter, we describe the basic principles of dynamic reflectometry that was used in



(a) (b) (c)

Fig. 1. (a) Input frame, (b) body model in same pose, and (c) silhouette matching.

the original pipeline in order to motivate where the novel algorithm described in Sect. 4 comes into play.

3.1 Acquisition

Inputs to our method are synchronized multi-view video sequences captured with eight calibrated cameras that feature 1004x1004 pixel image sensors and record at 25 fps. The cameras are placed in an approximately circular arrangement around the center of the scene which is illuminated by two calibrated spot lights. Since we conceptually separate the BRDF estimation from the estimation of the dynamic normal maps, we record two types of multi-view video sequence for each person and each type of apparel. In the first type of sequence, the so-called reflectance estimation sequence (RES), the person performs a simple rotation if front of the acquisition setup. One RES is recorded for each actor and each type of apparel, and it is later used to reconstruct the per-texel BRDF models. In the second type of sequence, the so-called dynamic scene sequence (DSS), the actor performs arbitrary movements. Several DSS are recorded, and from each of them, one relightable free-viewpoint video clip is reconstructed. Also the second component of our dynamic reflectance model, the dynamic normal maps, are reconstructed from each DSS.

3.2 Reconstructing Dynamic Human Shape and Motion

We employ an analysis-through-synthesis approach to capture both shape and motion of the actor from multi-view video footage without having to resort to optical markers in the scene. It employs a template human body model consisting of a kinematic skeleton and a single-skin triangle mesh surface geometry [2, 20]. In an initialization step, the shape and proportions of the template are matched to the recorded silhouettes of the actor. After shape initialization, the model is made to follow the motion of the actor over time by inferring optimal

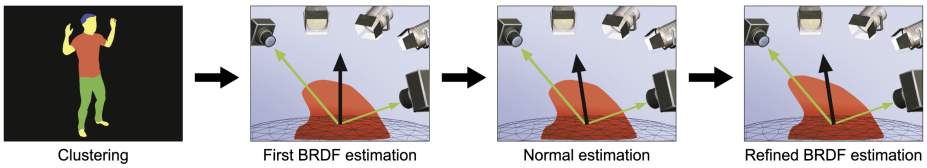


Fig. 2. Steps to estimate per-texel BRDFs.

pose parameters at each time step of video using the same silhouette matching principle, Fig. 1. We apply this dynamic shape reconstruction framework to every time step of each captured sequence, i.e. both RES and DSS. This way, we know for each time step of video the orientation of each surface point with respect to the acquisition setup which is a precondition for the subsequent dynamic reflectometry procedure.

3.3 Dynamic Reflectometry

Our dynamic reflectance model consists of two components, a static parametric isotropic BRDF for each surface point [16, 10], as well as a description of the time-varying direction of the normal at each surface location. The first component of the reflectance model is reconstructed from the video frames of the reflectance estimation sequence, the second component is reconstructed from each dynamic scene sequence. We formulate BRDF reconstruction as an energy minimization problem in the BRDF parameters [21]. This minimization problem has to be solved for each surface point separately.

The energy functional measures the error between the recorded reflectance samples of the point under consideration and the predicted surface appearance according to the current BRDF parameters. Given estimates of the BRDF parameters, we can also refine our knowledge about surface geometry by keeping the reflectance parameters fixed and minimizing the same functional in the normal direction, Fig 2. Once the BRDF parameters have been recovered from the RES, a similar minimization procedure is used to reconstruct the time-varying normal field from each DSS.

Before estimation commences, the surface model is parameterized over the plane and all video frames are transformed into textures. The estimation process is complicated by the fact that our shape model is only an approximation. Furthermore, potential shifting of the apparel over the body surface while the person is moving contradicts our assumption that we can statically assign material properties to individual surface points. We counter the first problem by means of an image-based warp correction step, and solve the latter problem by detecting and compensating textile shift in the texture domain. For details on each of these steps, please refer to [21].

In the original pipeline, as it was summarized above, we have estimated BRDF parameters for each surface point by taking only reflectance samples of

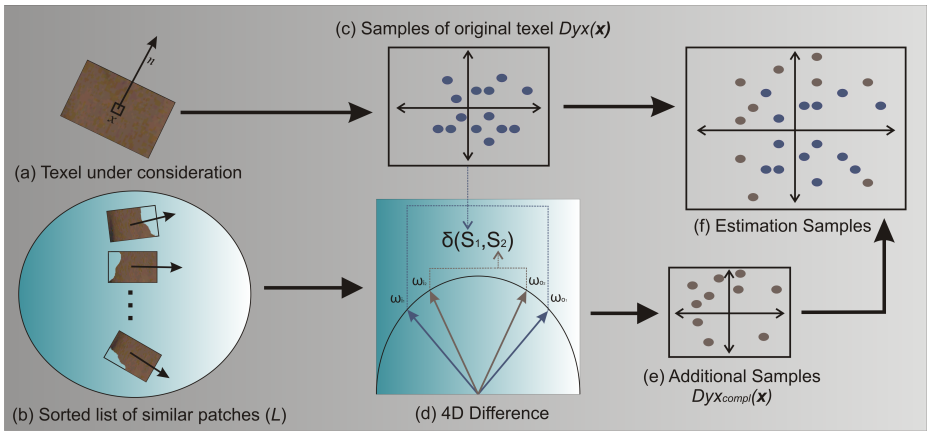


Fig. 3. Weighted selection of samples. Samples from the similar patches are added to the samples from the original texel. Additional samples are selected according to a weighting criteria that is based on their maximum angular difference from the samples of original texel.

this particular point itself into account [21]. In the following, we present a novel spatio-temporal sampling scheme that reduces the risks of a bias in the BRDF estimates by also taking into account dynamic reflectance samples from other surface points with similar material properties.

4 Spatio-temporal Reflectance Sharing

By looking at its appearance from each camera view over time, we can generate for each surface point, or equivalently, for each texel $\boldsymbol{x}$ a set of N appearance samples

$$\mathrm{Dyx}(\boldsymbol{x}) = \{S_i \mid S_i = (I_i, \hat{l}_i, \hat{v}_i), i \in \{1, \ldots, N\}\} \tag{1}$$

Each sample S_i stores a tuple of data comprising of the captured image intensity I_i (from one of the cameras), the direction to the light source $\hat{l}_i$, and the viewing direction $\hat{v}_i$. Please note that only if a point has been illuminated by exactly one light source, a sample is generated. If a point is totally in shadow, illuminated by two light sources, or not seen from the camera, no sample is created. Our acquisition setup comprising of only 8 cameras and 2 light sources is comparably simple and inexpensive. However, the fixed relative arrangement of cameras and light sources may induce a bias in $\mathrm{Dyx}(\boldsymbol{x})$. There are two primary reasons for this:

– Due to the fixed relative arrangement of cameras and light sources, each surface point is only seen under a fixed number of half vector directions $\hat{h} = \hat{l} + \hat{v}$.

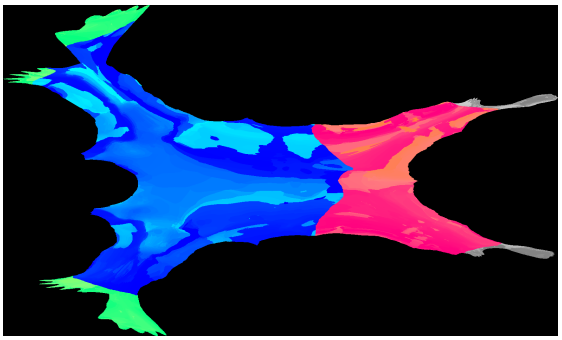


Fig. 4. Texture-space layout of surface patches. Patches of same material are clustered according to the average normal direction. For this illustration, patches of the same material are colored in the same overall tone (e.g. blue for the shirt) but different intensities.

– Even if the person performs a very expressive motion in the RES, samples lie on "slices" of the hemispherical space of possible incoming light and outgoing viewing directions.

Both of these factors possibly lead to BRDF estimates that may not generalize well to lighting conditions that are very different to the acquisition setup.

By means of a novel spatio-temporal sampling strategy, called spatio-temporal reflectance sharing, we can reduce the bias, Fig. 3. The guiding idea behind this novel scheme is to use more than the samples $\mathrm{Dyx}(\boldsymbol{x})$ that have been measured for the point $\boldsymbol{x}$ itself while the BRDF parameters for the point $\boldsymbol{x}$ are estimated. The additional samples, combined in a set $\mathrm{Dyx}_{\mathrm{compl}}(\boldsymbol{x})$, stem from other locations on the surface that are made of similar material. These additional samples have potentially been seen under different lighting and viewing directions than the samples from $\mathrm{Dyx}(\boldsymbol{x})$ and can thus expand the sampling range. It is the main challenge to incorporate these samples into the reflectance estimation at $\boldsymbol{x}$ in a way that augments the generality of the measured BRDFs but does not compromise the ability to capture spatial variation in surface appearance.

By explaining each step that is taken to draw samples for a particular surface point $\boldsymbol{x}$, we illustrate how we attack this challenge:

In a first step, the surface is clustered into patches of similar average normal directions and same material, Fig. 4. Materials are clustered by means of a simple k-means clustering using average diffuse colors [21]. The normal direction $\hat{n}$ of $\boldsymbol{x}$ defines the reference normal direction, Fig. 3a. Now, a list L of patches consisting of the same material as $\boldsymbol{x}$ is generated. L is sorted according to increasing angular deviation of average patch normal direction and reference normal direction, Fig. 3b. Now, n_p many patches $P_0, ..., P_{n_p}$ are drawn from L by choosing every lth list element. From each patch, a texel is selected at random, resulting in a set of texels, $T = \boldsymbol{x}_{P_0}, ..., \boldsymbol{x}_{P_{n_p}}$. The set of texels T has been

selected in a way that maximizes the number of different surface orientations. From the reflectance samples associated with texels in T, we now select a subset $\mathrm{Dyx}_{\mathrm{compl}}(\boldsymbol{x})$ that maximizes the coverage of the 4D hemispherical space of light and view directions. In order to decide which samples from T are potential candidates for this set, we employ the following selection mechanism.

A weighting function $\delta(S_1, S_2)$ is applied that measures the difference of two samples $S_1 = (\hat{l}_1, \hat{v}_1)$ and $S_2 = (\hat{l}_2, \hat{v}_2)$ in the 4D sample space as follows:

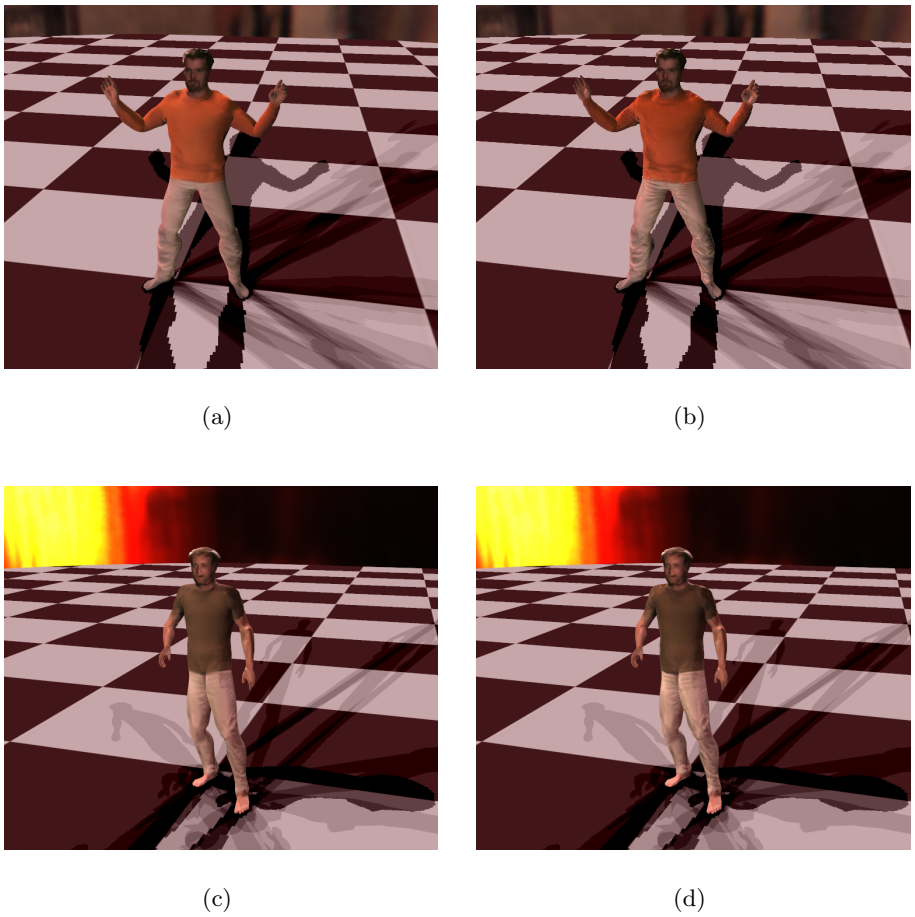
$$\delta(S_1, S_2) = \Delta(\hat{l}_1, \hat{l}_2) + \Delta(\hat{v}_1, \hat{v}_2) \tag{2}$$

where Δ denotes the angular difference between two vectors. We employ δ to select for each sample S_r in T its closest sample S_{closest} in $\mathrm{Dyx}(\boldsymbol{x})$, i.e. the sample for which $\omega_{S_r} = \delta(S_r, S_{\mathrm{closest}})$ is minimal, Fig. 3d. Each sample S_r is now weighted by ω_{S_r}. Only the $\lceil \alpha N \rceil$ samples from T with the highest weights eventually find their way into $\mathrm{Dyx}_{\mathrm{compl}}(\boldsymbol{x})$, Fig. 3e. The BRDF parameters for $\boldsymbol{x}$ are estimated by taking all of the samples from $\mathrm{Dyx}(\boldsymbol{x}) \bigcup \mathrm{Dyx}_{\mathrm{compl}}(\boldsymbol{x})$ into account, Fig. 3f. Through experiments we have found out that a value of $\alpha = 0.66$ represents a good compromise between estimation robustness and increase in computation time.

5 Results

We have tested our spatio-temporal reflectance sharing method on several input sequences. The sequences of 150-300 frames cover two different human subjects with different types of apparel. The method integrates seamlessly in the original pipeline and no modifications of any kind are required for the rendering system. We have verified both visually and quantitatively that our novel reflectance sampling method leads to BRDF estimation that generalizes better to lighting conditions different from the acquisition setup. Fig. 5 shows a side-by-side comparison between the results obtained with and without spatio-temporal reflectance sharing. Both human subjects are rendered under real world illumination using HDR environment maps. Importance sampling is used to obtain direction light sources that approximate the lighting from the static environment map [7]. Relightable free-viewpoint videos can be rendered from arbitrary viewpoints and under arbitrary lighting conditions at 6 fps if 16 approximating lights are employed. One can see that with the exploitation of spatial coherence, more surface detail is preserved under those lighting conditions which are strongly different from acquisition setup. A small comparison video demonstrating only the relighting can be seen here: http://www.mpi-inf.mpg.de/∼nahmed/Mirage2007.avi.

In addition to visual comparison, we also validated the method by comparing the average peak-signal-to-noise-ratio with respect to input video stream obtained under different lighting conditions. We have recorded one of our male test subjects under two different calibrated lighting setups, henceforth termed LC A and LC B. In each of the lighting setups, just one spot light has been employed. The positions of the light sources in LC A and LC B are (angularly) approximately 45° apart with respect to the center of the scene. We reconstructed the



(a) (b)

(c) (d)

Fig. 5. Comparison of renditions under captured real-world illumination such as the St Peter's Basilica environment map (a),(b) and the Grace Cathedral environment (c),(d) provided by Paul Debevec. One can see that compared to renditions obtained without spatio-temporal reflectance sharing ((a) (c)), subtle surface details are much better reproduced in the renditions obtained with spatio-temporal reflectance sharing ((b) (d)).

BRDF of the test subject under lighting setup LC B with and without our new reflectance sampling. Subsequently, we calculated the PSNR with the ground truth images of the person illuminated under setup LC A. Using our novel sampling method, we have estimated surface reflectance using different percentages of additional samples. For each case, we computed the PSNR with respect to the ground truth. Fig. 6 shows the results that we obtained. Note that the graph of

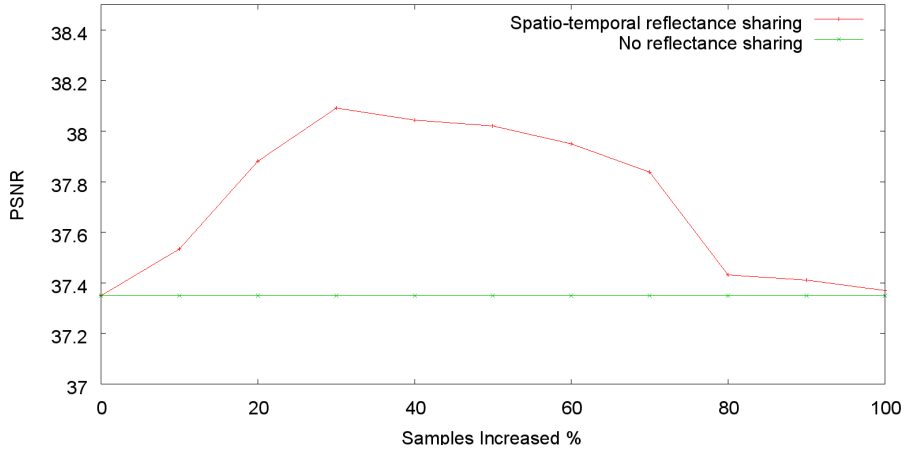


Fig. 6. PSNR values with respect to ground truth for different numbers of additional samples $\mathrm{Dyx}_{\mathrm{compl}}(\boldsymbol{x})$.

the original method (green line) is constant over the increasing number of samples just for the illustration purpose because it only considers the samples from a single texel. With spatio-temporal reflectance sharing (red line) both results are exactly the same in the beginning as no additional samples are considered, but it can be seen that the PSNR improves as additional samples are taken into account. We get a peak at around 30%-40% of additional samples. With the inclusion of more samples the PSNR gradually decreases as the ever increasing number of additional samples compromises the estimation of the reflectance's spatial variance. At maximum, we obtain a PSNR improvement of 0.75 dB. Although we have performed the PSNR evaluation only for one sequence, we are confident that for others it will exhibit similar results. This assumption is further supported by the more compelling visual appearance obtained for all the other test data that we have used.

Our approach is subject to a couple of limitations. We currently neglect interreflections on the body. In the RES, they potentially play a role between the wrinkles in clothing. For reflectance sharing, samples from the wrinkles can lead to erroneous estimation. To prevent this effect from degrading the estimation accuracy, we have taken care to minimize the number of wrinkles in the RES. Sometimes, we observe small rendering artefacts due to undersampling (e.g. on the underneath of the arms). However, we have verified that the application of a RES sequence showing several rotation motions with different body postures almost completely solves this problem. If even in this extended RES a pixel is never seen by any of the cameras, we fill in reflectance properties from neighboring regions in texture space.

Despite the limitations, our results show that spatio-temporal reflectance sharing enable faithful estimation of dynamic reflectance models with only a handful of cameras.

6 Conclusions

We have presented a spatio-temporal reflectance sharing method that reduces the bias in BRDF estimation for dynamic scenes. Our algorithm exploits spatial coherence by pooling samples of different surface location to robustify reflectance estimation. In addition, it exploits temporal coherence by taking into consideration samples from different steps of video. Despite the spatial-temporal resampling, our algorithm is capable of reliably capturing spatially-varying reflectance properties. By means of spatio-temporal reflectance sharing, we obtain convincing 3D video renditions in real-time even under lighting conditions which differ strongly from the acquisition setup.

7 Acknowledgements

This project has been supported by EU 3DTV NoE project No 511568.

References

1. Fausto Bernardini, Ioana M. Martin, and Holly Rushmeier. High-quality texture reconstruction from multiple scans. *IEEE TVCG*, 7(4):318–332, 2001.
2. J. Carranza, C. Theobalt, M.A. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. In *Proc. of SIGGRAPH'03*, pages 569–577, 2003.
3. Per Einarsson, Charles-Felix Chabert, Andrew Jones, Wan-Chun Ma, Bruce Lamond, im Hawkins, Mark Bolas, Sebastian Sylwan, and Paul Debevec. Relighting human locomotion with flowed reflectance fields. In *Rendering Techniques*, pages 183–194, 2006.
4. Athinodoros S. Georghiades. Recovering 3-d shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering*, pages 230–240, 2003.
5. D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Proc. of ICCV*, pages 341–448, 2004.
6. Markus H. Gross, Stephan Würmlin, Martin Näf, Edouard Lamboray, Christian P. Spagno, Andreas M. Kunz, Esther Koller-Meier, Tomás Svoboda, Luc J. Van Gool, Silke Lang, Kai Strehlke, Andrew Vande Moere, and Oliver G. Staadt. blue-c: a spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph. (Proc. of SIGGRAPH'03)*, 22(3):819–827, 2003.
7. Vlastimil Havran, Miloslaw Smyk, Grzegorz Krawczyk, Karol Myszkowski, and Hans-Peter Seidel. Importance Sampling for Video Environment Maps. In *ACM SIGGRAPH 2005 Full Conference DVD-ROM*, Los Angeles, USA, August 2005. ACM SIGGRAPH, ACM. Sketches & Applications.
8. T. Hawkins, A. Wenger, C. Tchou, A. Gardner, F. Göransson, and P. Debevec. Animatable facial reflectance fields. In *Proc. of Eurographics Symposium on Rendering*, pages 309–319, 2004.

9. Takeo Kanade, Peter Rander, and P. J. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE MultiMedia*, 4(1):34–47, 1997.
10. E. Lafortune, S. Foo, K. Torrance, and D. Greenberg. "non-linear approximation of reflectance functions". In *Proceedings of SIGGRAPH'97*, pages 117–126, August 1997.
11. Hendrik P. A. Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*, 22(2):27, 2003.
12. Ming Li, Hartmut Schirmacher, Marcus Magnor, and Hans-Peter Seidel. Combining stereo and visual hull information for on-line reconstruction and rendering of dynamic scenes. In *Proc. of IEEE Multimedia and Signal Processing*, pages 9–12, 2002.
13. T. Matsuyama and T. Takai. Generation, visualization, and editing of 3D video. In *Proc. of 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)*, page 234ff, 2002.
14. W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of ACM SIGGRAPH 00*, pages 369–374, 2000.
15. W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. Graph. (Proc. SIGGRAPH'03)*, 22(3):759–769, 2003.
16. B.-T. Phong. Illumnation for computer generated pictures. *Communications of the ACM*, pages 311–317, 1975.
17. H. Rushmeier, G. Taubin, and A. Guéziec. Applying Shape from Lighting Variation to Bump Map Capture. In *Eurographics Workshop on Rendering*, pages 35–44, June 1997.
18. Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. Object Shape and Reflectance Modeling from Observation. In *Proc. of SIGGRAPH'97*, pages 379–388, 1997.
19. J. Starck, G. Miller, and A. Hilton. Video-based character animation. In *Proc. ACM Symposium on Computer Animation*, 2005.
20. C. Theobalt, J. Carranza, M. Magnor, and H.-P. Seidel. Combining 3d flow fields with silhouette-based human motion capture for immersive video. *Graphical Models*, 66:333–351, September 2004.
21. Christian Theobalt, Naveed Ahmed, Hendrik P. A. Lensch, Marcus Magnor, and Hans-Peter Seidel. Seeing people in different light - joint shape, motion and reflectance capture. *IEEE Transactions on Visualization and Computer Graphics*, to appear.
22. M. Waschbüsch, S. Würmlin, D. Cotting, F. Sadlo, and M. Gross. Scalable 3D video of dynamic scenes. In *Proc. of Pacific Graphics*, pages 629–638, 2005.
23. A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. In *ACM TOG (Proc. of SIGGRAPH'05)*, volume 24(3), pages 756–764, 2005.
24. S. Würmlin, E. Lamboray, O.G. Staadt, and M.H. Gross. 3d video recorder. In *Proc. of IEEE Pacific Graphics*, pages 325–334, 2002.
25. Ruo Zhang, Ping-Sing Tsai, James Cryer, and Mubarak Shah. Shape from Shading: A Survey. *IEEE Trans. PAMI*, 21(8):690–706, 1999.
26. Todd Zickler, Sebastian Enrique, Ravi Ramamoorthi, and Peter N. Belhumeur. Reflectance sharing: Image-based rendering from a sparse set of images. In *Proc. of Eurographics Symposium on Rendering*, pages 253–264, 2005.
27. C. Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. High-quality video view interpolation using a layered representation. *ACM TOC (Proc. SIGGRAPH'04)*, 23(3):600–608, 2004.