Citation for published version:
Tarko, J & Richardt, C 2018, 'Environment Reflections with 360° Videos using Omnidirectional Structure from Motion' Paper presented at CVMP 2018, London, UK United Kingdom, 13/12/18 - 14/12/18, .

Publication date:
2018

Document Version
Peer reviewed version

Link to publication

University of Bath

Environment Reflections with 360° Videos using Omnidirectional Structure from Motion

Joanna Tarko[1, 2]
J.K.Tarko@bath.ac.uk
Christian Richardt[1]
christian@richardt.name

[1] Department of Computer Science, University of Bath

[2] Checkmate VR / DC Activ

Image-based lighting (IBL) reproduces lighting conditions captured in an image using complementary environment maps acquired from light probes. Environment maps are then used to correctly light computer-generated (CG) elements that are inserted into the image. They also provide a source of reflections for reflective materials [1].

Originally, light probes were mirror balls that have to be photographed and then the photographs were converted into environment maps. Sometimes fisheye cameras or panoramic images were used instead. However, in recent years, affordable 360° spherical cameras have become widely available. They were soon adopted into image-based lighting pipelines as they provide easily obtainable environment maps, not only for static environments but also for temporally changing ones. For mixed-reality rendering, 360° video can therefore act simultaneously as background footage and temporally varying environment map [3].

However, a single environment map is not sufficient to obtain spatially correct reflections. Michiels et al. [4] combined omnidirectional structure from motion (SfM) with mixed-reality rendering to render CG elements that are inserted into 360° video footage according to their position relative the original camera path, so the closest frame could be used as an environment map. We propose an approach that extends theirs with the following contributions:

- We employ inverse tone mapping, which provides high dynamic range (HDR) environment maps that more accurately reproduce real-world lighting conditions.
- Our own implementation of omnidirectional structure from motion, which performs tracking directly on stitched equirectangular video. This format is currently the most popular for 360° video, as most cameras perform stitching internally.
- We integrate our approach with the Unity game engine for mixed-reality rendering of objects into the video with real-time dynamic lighting while enabling user interaction.

We first compute omnidirectional structure from motion to reconstruct the camera path and a sparse 3D point cloud to guide the insertion of CG elements. Specifically, we use a KLT tracker that we modified to work on equirectangular video to find point correspondences between frames. We use forward and backward tracking to maximise the length of feature trajectories, and apply the epipolar constraint for spherical cameras to calculate relative camera poses. We triangulate the 3D positions of our sparse reconstruction and then refine both the initial camera poses and reconstructed points with hierarchical bundle adjustment using Ceres.

Secondly, we stabilise the 360° video by rotating every video frame in the opposite direction of its camera orientation. This aligns all environment maps with the axes of the global coordinate system and ensures that lighting and reflections are correctly oriented in Unity.

Then, we apply inverse tone mapping [2] to the stabilised 360° video to obtain HDR environment maps. We import these into Unity together with the reconstructed camera path and 3D points from the first step. Our script takes the camera path and divides it into a number of segments specified by the user, placing a reflection probe in each segment. Reflection probes are a Unity feature, which acts as a local environment map for reflections, and reflective objects are affected by the nearest probe.

We tested our approach on a video recorded with the Ricoh R 360° video camera. In Figure 1, we placed CG mirror spheres in the scene at different distances from the camera and compared their reflections produced by two types of environment maps. With a single global environment map centred in the beginning of the world coordinate system, reflections do not show the correct perspective. We thus use a set of local reflection maps, which make reflections more convincing as they take into account the position from the nearest reflection probe. This produces the most accurate results for objects placed as close to the original camera path as possible.

In future work, we plan to use our current sparse scene reconstruction as a guidance for a simple primitive-based reconstruction, such as a ground plane and walls. This introduces additional interactions between CG elements and the captured real scene, for example by casting virtual shadows onto real objects.

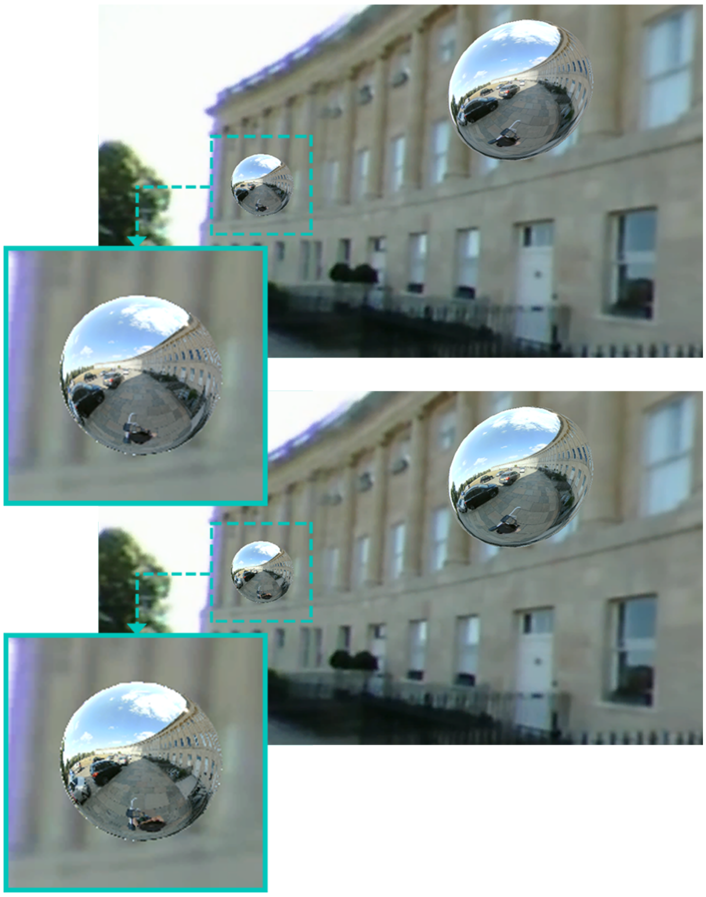

Figure 1: Computer-generated mirror spheres are inserted at two different locations in the scene. **Top:** using a single global environment map results in identical reflections in both mirror spheres, which is incorrect (compare the position of the black car on the left-hand side of the sphere between the near and far spheres). **Bottom:** using frames along the camera path as environment maps produces visibly different reflections (see the black car on the left and the house entrance on the right).

This research was funded by the Centre for Digital Entertainment (EPSRC CDT EP/L016540/1), and by Checkmate VR and DC Activ as part of their R&D programme.

[1] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, pages 189–198, 1998.

[2] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 36(6):177:1–10, November 2017.

[3] Thomas Iorns and Taehyun Rhee. Real-time image based lighting for 360-degree panoramic video. In *Proc. Pacific-Rim Symposium on Image and Video Technology (PSIVT) Workshops*, volume 9555 of *LNCS*, pages 139–151, 2016.

[4] Nick Michiels, Lode Jorissen, Jeroen Put, and Philippe Bekaert. Interactive augmented omnidirectional video with realistic lighting. In *Augmented and Virtual Reality*, volume 8853 of *LNCS*, pages 247–263, 2014.