Voxelizing Light-Field Recordings

David C. Schedl, Indrajit Kurmi, Oliver Bimber

Institute of Computer Graphics, Johannes Kepler University Linz, Austria

Contact: firstname.lastname@jku.at

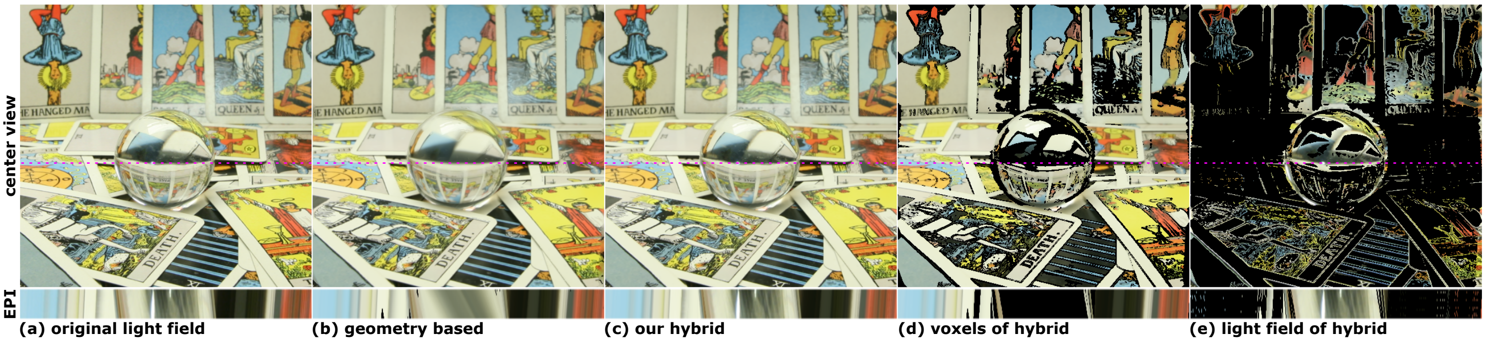

Figure 1: Comparison of a geometry-based representation (b) to our hybrid (c) of light-field and voxelized light-field data. The original light field (a) consumes 216.8MB of memory (8bit per channel; 3 colors), while our hybrid representation compresses the data to 82.4MB (38%). Isotropic and uniform scene areas can be represented as voxels (color+depth), requiring only 0.5MB of sparse data, without a perceivable loss of quality (d). Anisotropic scene parts (e.g. reflections and refractions) and occlusion boundaries (producing errors in depth reconstruction) are represented as sparse light field, requiring 81.9MB of data (e). The epipolar images (EPI) clearly show that anisotropic scene (e.g., the specular highlight) areas are retained in our hybrid, while they are completely lost in the geometry-based representation.

Introduction

Light fields are an emerging image-based technique that support free viewpoint navigation of recorded scenes as demanded in several recent applications, such as Virtual Reality. Typically, game engines or 360° images provide content for Virtual Reality applications. While real-time renderings of game assets lack the visual fidelity of recorded scenes, monoscopic or stereoscopic 180/360° recordings (e.g., Facebook's 3D-180° video [1]) lack the necessary motion parallax and do not store any view-dependent effects (e.g., the change of reflections and refractions as the viewer moves). Light fields are suitable for capturing and rendering such complex effects. Pure image-based representations, however quickly become inefficient, as many images are required to be captured, stored, and processed. Geometric scene representations require less storage and are more efficient to render. Geometry reconstruction, however, is unreliable and might fail for complex scene parts. Furthermore, view-dependent effects that are preserved with light fields are lost in pure geometry-based techniques. Therefore, we propose a hybrid representation and rendering scheme for recorded dense light fields: we extract isotropic scene regions and represent them by voxels, while the remaining areas are represented as sparse light field. In comparison to dense light fields, storage demands are reduced while visual quality is sustained.

Our Method

Our algorithm takes a densely sampled light field as input and extracts isotropic scene regions that can be represented by color and depth into a voxel representation, while the remaining areas are represented as sparse light field. We start by rendering the light field with different focal depths from front to back and store the result in a focal stack. Additionally, we compute a directional consistency value (i.e., the color variance over all directions) for every point on every slice of the focal stack, as described in [2]. The directional variance will be high in out-of-focus areas, at occlusion boundaries, and at anisotropic regions (cf. Fig. 2). For every pixel we determine the layer of the focal stack where the corresponding (orthographic projected) entry has minimum directional variance. This is done to remove the contributions of out-of-focus areas. All these entries are then used as voxels in our hybrid representation if the minimum directional variance is less than a predefined threshold. These are rendered into the pixels that are likely to be in focus and represent isotropic scene areas, as shown in Fig. 1(d) and Fig. 2(a). Next, we remove the rays from the original light field that intersect at these voxels. The remaining light-field rays represent anisotropic scene areas and occlusion boundaries, as shown in Fig. 1(e) and Fig. 2(b-d). We render the sparse voxel and light-field representations with the same camera parameters and blend the results as illustrated in Fig. 1(c).

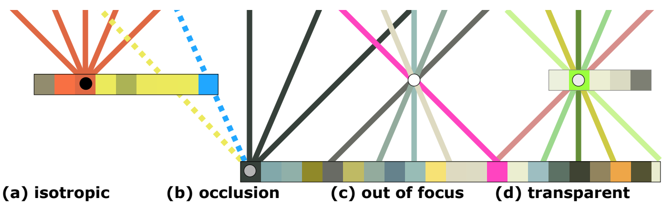

Figure 2: A simplified scene with 3 surfaces showing ray diagrams when light-field rendering is applied. If a non-occluded isotropic surface is rendered in focus, the color variance is low (a). In cases of occlusions (i.e., some rays originate from occluders in the foreground), the color consistency value is high (b). Occluded rays are indicated by dotted lines. Out-of-focus areas result in high variance as rays originate from different regions of the scene (c). The color consistency value is high for anisotropic scene regions (i.e., transparency) even if they are rendered in focus, because ray colors change for varying views (d).

Result and Outlook

Figure 1 and 3 shows renderings of our hybrid technique for the Tarot [3] light field of size 512x512x17x17 rays. Our algorithm is implemented in Matlab, we used 50 slices for the focal stack, and a variance threshold of 7.7e-04 for each color channel ranging from 0 to 1. Our current light-field voxelization approach has several shortcomings: First, variance thresholding must be set manually and is not robust against miscalibrations in the recorded light-field data. Perceptual metrics that allow a comparison of intermediate results with the original light field might enable for automatic thresholding. Second, we currently do not use optimized data structures for sparse data representations. While they exist for sparse volume data sets, we need to investigate efficient data structures for sparse light fields that support compressed storage as well as efficient reading and rendering. Furthermore, we need to develop real-time rendering techniques for consistent visualization of both scene components. In comparison to classical light-field rendering, we expect vast improvements in rendering speed, due to the decreased computational demand mainly caused by the reduced light-field size.

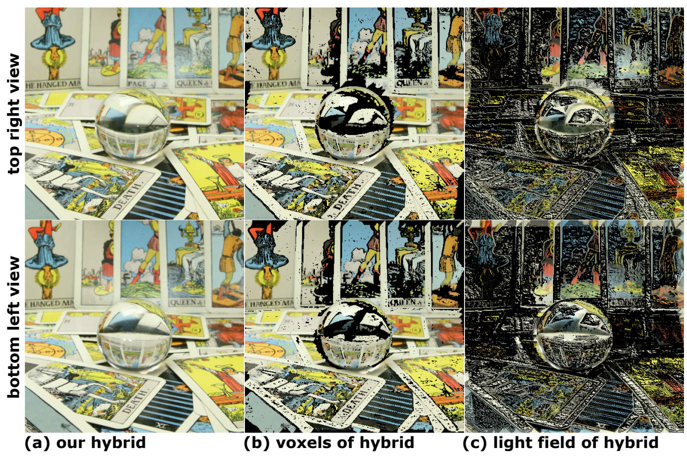

Figure 3: Hybrid rendering (a) of light-field and voxelized light-field data for extreme perspectives of Fig. 1. Anisotropic scene parts (e.g., reflections and refractions on the glass sphere and on the cards in the back) and occlusion boundaries are represented as sparse light field (c). The remaining isotropic and uniform scene areas are represented as voxels (b). Further views are shown in the Supplementary Video; online at: https://youtu.be/tL7UEmGoJ8k.

video URL

References

[1] Facebook 3D-180 Video, https://facebook360.fb.com/2018/06/15/introducing-3d-180-video-on-facebook/
[2] Zhang C., Chen T.: A self-reconfigurable camera array. In ACM SIGGRAPH Sketches (2004), p. 151. 2
[3] The (New) Stanford Light Field Archive, Stanford University, http://lightfield.stanford.edu/lfs.html

JOHANNES KEPLER UNIVERSITY LINZ

EG2019