Compositing Light field Video Using Multiplane Images

Matthew DuVall
matthewduvall@google.com
Google, Inc.

John Flynn
jflynn@google.com
Google, Inc.

Michael Broxton
broxton@google.com
Google, Inc.

Paul Debevec
debevec@google.com
Google, Inc.

Figure 1: (left) Light field camera generated multiplane images (middle) Composites with segmentated portions

ABSTRACT

We present a variety of new compositing techniques using Multiplane Images (MPI's) [Zhou et al. 2018] derived from footage shot with an inexpensive and portable light field video camera array. The effects include camera stabilization, foreground object removal, synthetic depth of field, and deep compositing. Traditional compositing is based around layering RGBA images to visually integrate elements into the same scene, and often requires manual 2D and/or 3D artist intervention to achieve realism in the presence of volumetric effects such as smoke or splashing water. We leverage the newly introduced DeepView solver [Flynn et al. 2019] and a light field camera array to generate MPIs stored in the DeepEXR format for compositing with realistic spatial integration and a simple workflow which offers new creative capabilities. We demonstrate using this technique by combining footage that would otherwise be very challenging and time intensive to achieve when using traditional techniques, with minimal artist intervention.

CCS CONCEPTS

• **Computing methodologies → Image processing**.

KEYWORDS

light field, compositing, volumetric imaging, view synthesis

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '19 Posters, July 28 - August 01, 2019, Los Angeles, CA, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6314-3/19/07.
https://doi.org/10.1145/3306214.3338614

ACM Reference Format:
Matthew DuVall, John Flynn, Michael Broxton, and Paul Debevec. 2019. Compositing Light field Video Using Multiplane Images. In *Proceedings of SIGGRAPH '19 Posters.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3306214.3338614

1 INTRODUCTION

Multiplane images encode three dimensional scene information as a series of front-parallel textured planes that fill a viewing frustum projected into a 3D environment. With an MPI it is trivial to render novel viewpoints simply by moving the virtual camera viewpoint relative to these planes. As such, MPIs can be utilized in most 3D rendering engines. The number of these planes determines the valid volume size, and is directly proportional to the physical dimensions of the capture rig array, as well as near and far content of the scene represented. Each plane not only displays a slice of the diffuse volumetric image, but also encodes view dependant effects such as reflections through the use of alpha compositing in back to front rendering. Occlusions, lighting reflections, thin structures, and scenes with high depth complexity can be realistically rendered. In this work we propose a Nuke based workflow and show examples that exploit the unique features of MPIs for generating 2D composites of 3D volumetric scene data.

To generate an MPI video, we capture image sequences of photography using a light field camera array. The array in our examples is 16 GoPro Hero 4 cameras capturing at a resolution of 2704 x 2028, at 29.97fps. The resulting mpeg4 data is processed through the DeepView solver. We note that a MPI derived using DeepView shares several of the advantages of computer generated deep image composites [Lokovic and Veach 2000] with the streamlined abilities within Nuke that are possible when using DeepEXRs [Kainz 2013]. Our DeepView MPIs were generated at 1920 x 1160 resolution projected onto 80 or 96 planes with a near depth of .5m and a far depth

of 100m. The first and last planes match the processed near and far parameters, with the number of planes dispersed between using inverse depth disparity, scaled to match the projected view frustum.

Viewing, modifying and compressing the MPI to the final 2D output starts with reassembling the representation within the desired 3D package. We used Nuke v11.2v5 for generating the necessary planes and texturing with the accompanying images. Each plane is spatially static in relation to the camera, but temporally dependent on the MPI image sequence. This method was chosen for the benefit of visualization and interaction in 3D and accessibility to Nuke 3D processing tools. Color corrections are applied post-composite to retain per plane RGBA value intensities and therefore MPI synthesis accuracy.

1.1 VIRTUAL CAMERA EDITING

MPIs effectively encode a light field representation that supports high quality view synthesis at render-time. This capability permits us to perform several effects during post-processing that otherwise would need to be captured in the original camera footage, or painstakingly created manually by an artist.

For example, the MPI view synthesis volume is large enough to allow camera movement within a range of just over half a meter. This permits us to apply motion correction to achieve a smooth camera move when working with hand held footage. By rendering the MPI to new rectilinear virtual camera and using existing camera tracking tools, we can derive and correct for scene motion. Rendering from a new, motion-stabilized camera trajectory while within the MPI viewing volume generates imagery that appears to have been captured with a steady-cam, as seen in our video example "Camera Stabilization." Alternate post production methods to generate a similar effect would require segmentation and re-projection onto tracked geometry, and it would result in stretched reflections across geometry surfaces. This method avoids such artifacts and substantially decreases the complexity of a camera stabilization workflow.

Another advantage of MPIs is that they can be processed to produce a view-dependent "composite" depth map. The composite depth map is generated by rendering the distance of each plane to the texture value premultiplied by the MPI alpha before compositing all planes together. The result is a high resolution, view-interpolated map suitable for effects such as depth of field simulation. The ability to display view dependent effects such as reflections and refractions translates through this composite depth map by describing the distance from camera to visible feature, rather than camera to geometry surface. As such surfaces such as water can be rendered more accurately with these effects than with traditional geometry depth maps.

A common challenge in post production is matching footage from multiple camera sources. Adjustments to focal length, lens distortion, depth of field, image sensor artifacts can be easily rendered using a virtual camera view of an MPI scene. For example, our supplemental video example "Virtual Camera Editing" shows a virtual camera that was stabilized and rendered to DeepEXR with a channel for the composite depth. Then, depth of field was simulated with ZDefocus node followed by lens distortions (radial and chromatic).

2 COMPOSITING AND SEGMENTATION

Within the MPI objects are segmented along the Z axis from camera. To segment an object, we must therefore isolate the planes (or portions thereof) containing portions of that object in their RGBA textures. This aligns well with common production techniques with a single loose rotoscoped matte projected from the center of the volume through the planes is often sufficient to isolate objects. Restricting the matte projection within a specified range of depth representing the object and its desired surface effects completes the segmentation. In our video example "Segmentation," the animated rotoscoped matte was a simple square around the fence, restricted between .5m and 1.0m.

Several considerations needed to achieve a successful segmentation. Varying the input detail and requested output detail have dramatic effects on the ability to segment. Objects closer to the camera segment from the background more effectively because they demonstrate more parallax in the input and pass through more planes on the output.

Rather than segmenting portions of the image, the capture as a whole can be merged with another. As MPI assets, multiple captures can be accurately rendered at once by triangulating the position of each pixel to determine the per pixel compositing order of all MPIs within the scene. In our Nuke 3D environment, this is simplified to merging MPI scenes before rendering. In our video example "DeepEXR Compositing," the foreground MPI and background MPI were rendered to DeepEXR for compositing. Because the scenes were so similar, segmentation and determination of foreground or background was not necessary and they were merged as whole DeepEXR assets. Differences did exist in the lighting between scenes, but the output image resolved without the introduction of artifacts.

3 CONCLUSION

We have presented a new method for compositing and manipulating video sequences in post-production using multiplane images in a popular commercial compositing tool, Nuke. We showed that MPIs generated with the DeepView solver can be used to achieve complex volumetric compositing and view synthesis effects without requiring substantial time and attention of a post production artist. Our approach shows the promise of light field video manipulation using MPIs for enhancing such post-production workflows.

REFERENCES

John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. 2019. DeepView: View synthesis with learned gradient descent. *CVPR* (2019).

Florian Kainz. 2013. Interpreting OpenEXR Deep Pixels. (2013). Retrieved from https://lists.nongnu.org/archive/html/openexr-devel/2013-09/pdfPyDOYBTkje.pdf.

Tom Lokovic and Eric Veach. 2000. Deep shadow maps. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (2000). https://doi.org/10.1145/344779.344958

Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. 2018. Stereo Magnification: Learning view synthesis using multiplane images. *Siggraph* (2018).