Deep Thoughts on Deep Image Compression

Rob Pieké
MPC Shadow Lab
rob-p@moving-picture.com

Yanli Zhao
MPC Shadow Lab
yanli-z@moving-picture.com

Fabià Serra Arrizabalaga
MPC Shadow Lab
fabia-sa@moving-picture.com

Figure 1: Deep image composite of seven furball elements before and after compression - image-set on the right is 75% smaller

ABSTRACT

Deep image compositing has, in the last decade, become an industry standard approach to combining multiple computer-generated elements into a final frame. With rich support for multiple depth-specified samples per-pixel, deep images overcome many of the challenges previously faced when trying to combine multiple images using simple alpha channels and/or depth values.

A practical challenge when using deep images, however, is managing the data footprint. The visual fidelity of the computer generated environments, characters and effects is continually growing, typically resulting in both a higher number of elements and greater complexity within each element. It is not uncommon to be using "gigabytes" to describe the size of deep image collections which, as more and more visual effects facilities establish a global presence, introduces a significant concern about timely overseas data transfer. Further, as deep images flow through compositing networks, the high sample count contributes to longer processing times.

Our observation is that, with a richer contextual understanding of the target composite, systems - both automatic and artist-controlled - can be built to significantly compress deep images such that there is no perceptual difference in the final result.

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 '18 Talks, August 12-16, 2018, Vancouver, BC, Canada
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5820-0/18/08.
https://doi.org/10.1145/3214745.3214753

CCS CONCEPTS

• **Computing methodologies → Image processing**; **Image compression**;

KEYWORDS

deep images, compositing, compression

ACM Reference Format:
Rob Pieké, Yanli Zhao, and Fabià Serra Arrizabalaga. 2018. Deep Thoughts on Deep Image Compression. In *Proceedings of SIGGRAPH '18 Talks.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3214745.3214753

1 CONTEXT

The investigations and results presented are grounded in the real-world workflows and challenges of MPC's rendering and compositing artists. Specifically, we have focused on deep OpenEXR content generated by Pixar's RenderMan and processed by Foundry's Nuke.

Inspecting typical VFX production deep images, we often observe a large number of samples per-pixel. This is particularly prevalent for images of furry creatures, and volumetric effects (mist, dust, clouds, etc) where many layers of fine sub-pixel and/or partially-transparent geometry contribute to the final appearance. While the data may accurately represent the scene, we believe it is often overkill in the final context and can be compressed without a visual impact.

We draw inspiration from the original paper on deep images [Lokovic and Veach 2000], where the authors describe a method to remove and modify samples from a piecewise-linear curve, such that it maintains a measurably-similar shape. Rather than considering a value error threshold for each sample, however, we instead attempt to define depth threshold metrics to allow clustering and merging of

nearby samples. Such "pre-compositing" can either collapse several point samples into a single point sample, or convert them into a single volume sample. We refer the reader to [Kainz 2013] for more details on the distinction between these sample types.

2 DISTANCE

Our ability to perceive the geometric and/or textural fidelity of an object is directly linked to our distance to said object. Whereas we can see every hair and freckle on our own hands in front of us, we cannot resolve the same level of detail on someone standing a few yards away. This has been exploited with great success in both VFX and game production, from distance-based geometry LODs [Clark 1976], mip-map textures [Williams 1983] and, most relevant to our work, voxelizing fur [Maurer et al. 2008].

Considering a single pixel in a deep image, we associate each sample with a distance-merging threshold based on the sample's z-distance from the camera, $d(z) = \alpha z^{\beta}$ (where α and β are artist-controllable values - though we typically leave β at a value of 1). For a given sample i with corresponding depth z_i, we identify all subsequent samples such that for sample j, $z_j - z_i \leq d(z_i)$. We then merge that group into a single point/volume sample. In Figure 2 below, the pink samples cannot be merged, whereas the blue samples can.

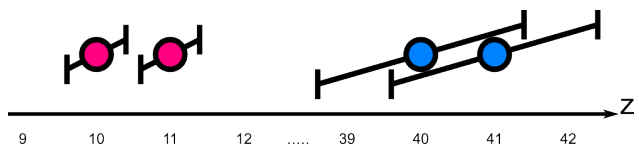

Figure 2: Four samples from a single deep pixel, and their corresponding $d(z_i)$ illustrated as ranges

3 DEPTH OF FIELD

Having observed significant file size reductions when leveraging the above metric, we gave thought as to what other situations could lead to a reduced ability to perceive detail and strict ordering of elements. At MPC, we generally perform depth of field as a post-process, and it seemed sensible that in heavily-blurred regions, we could likely sacrifice accuracy without a meaningful visual impact.

Taking a very similar approach as before, we note that samples with similar depth will blur with a similar radius, and samples farther from the point of focus will blur more than samples at or near it. With knowledge of the camera we want to emulate (focal length, focus distance, f-stop, etc.) we calculate the circle of confusion $c(z_i)$ for each sample i, and again iterate and merge where $z_j - z_i \leq \alpha c(z_i)^{\beta}$.

4 ELEMENT OVERLAP

Ultimately, deep data only provides us value if we actually perform depth-based calculations, where compositing multiple elements together is by far the most common. If we have no other need for the depth information then we can perform an inexpensive post-process to consider where the elements overlap. Starting with the 2D bounding boxes, any non-overlapping regions can have all their pixels flattened. For overlapping regions, we can iterate over the pixels and look at the Z ranges for contributing elements. Again if there are no overlaps, the pixels can be flattened. By only considering bounding regions, we can process the image set very quickly, similar to broadphase collision detection in rigid body simulation.

5 AVOIDING ARTIFACTS

As we began to apply the above techniques on production data, we ultimately found them to be too naïve. Much as was done in [Lokovic and Veach 2000], it proved important for us to consider the actual sample values, not only their depths. A very dark sample behind a very bright sample suggests there is likely an in-between blocking/shadowing sample from another not-yet-considered element. If we merge the two samples, the dark background will "bleed" into the foreground, objectionably dimming highlights in the final composite. To combat this, we introduced a limiting threshold for merging based on the relative luminance of successive samples.

6 IMPACT AND FURTHER THOUGHTS

For our production content, we're observing the largest impact on elements with furry creatures or volumetric effects, which is particularly fortunate as those have historically been, by far, our largest deep image files. While we don't yet have a full production's worth of data to analyze, early findings suggest we can reduce most elements of this type by 50-90% in size without objectionable artifacts. With these smaller files, we also observe our image-based depth of field plugin running 2-3x faster than before.

We feel that we've only scratched the surface of possible metrics that can be used to aid compression, and are still investigating how they can be combined and applied at different points in MPC's VFX process. We are identifying conservative but robust settings that can be used to automatically post-process images directly coming out of the renderer, and providing the artists with tools to apply more aggressive compression where it makes sense.

ACKNOWLEDGMENTS

Special thanks go to Kirstin Hall for providing endless practical insight into MPC's compositing workflows, and to Kai Wolter for taking our work out of the lab and into production.

REFERENCES

James H. Clark. 1976. Hierarchical Geometric Models for Visible Surface Algorithms. *Commun. ACM* 19, 10 (Oct. 1976), 547–554. https://doi.org/10.1145/360349.360354

Florian Kainz. 2013. Interpreting OpenEXR Deep Pixels. http://www.openexr.com/InterpretingDeepPixels.pdf. (2013).

Tom Lokovic and Eric Veach. 2000. Deep Shadow Maps. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 385–392. https://doi.org/10.1145/344779.344958

Eric Maurer, Sean Palmer, Alen Lai, and Jamie Williams. 2008. Countless Characters and Clovers: Interpreting Dr. Seuss' Style with 3D Fur. In *ACM SIGGRAPH 2008 Talks (SIGGRAPH '08)*. ACM, New York, NY, USA, Article 63, 1 pages. https://doi.org/10.1145/1401032.1401113

Lance Williams. 1983. Pyramidal Parametrics. In *Proceedings of the 10th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '83)*. ACM, New York, NY, USA, 1–11. https://doi.org/10.1145/800059.801126