Real-Time Realistic Pixel Synthesis using Deep Learning for Augmented and Virtual Reality

PI: Michael Doggett, Department of Computer Science, Lund University
co-PI: Patric Ljung, Department of Science and Technology, Linköping University

1 Executive summary

Augmented and Virtual Reality will have a major impact on future Human Computing Interfaces. This project aims to solve the challenges of generating the realistic images necessary to ensure the level of immersion to make these new platforms essential. By using high quality physical accurate resources, combined with the latest techniques in Deep Learning combined with Real-Time Rendering, we will address the challenges of immersive realistic imagery for Augmented and Virtual Reality.

2 Research area, motivation, and research challenges

Augmented and Virtual Reality (also referred to as XR) have advanced significantly in recent years, with massive industrial investment into the next generation of devices across all major technology companies, including Microsoft, Facebook, Google and Apple. The availability of these display technologies, and recent developments in graphics hardware lays the foundations to create visual experiences beyond what is possible today. To make these technologies into the next computing platform, many challenges remain to create the necessary visual experience. For immersive visual experiences, high resolutions at high frame rates are essential. Solving these problems will allow XR to have a transformative impact on the future of human computer interaction.

To address these challenges we focus on two research questions:

RQ1 *How can graphics hardware be used to generate billions of pixels per second for future XR technologies?*

RQ2 *How can real-world scenes and objects, be faithfully represented with reasonable amounts of captured data and enhanced using procedural details to create realistic scene representations?*

Human Visual Perception can sense 400 million pixels at frame rates around 240Hz. Future micro LED displays will soon be able to deliver these resolutions close to the human eye enabling new experiences in XR. To fill these displays with high quality pixels will require Real-Time Rendering algorithms with realism and performance far beyond the capabilities of current graphics software algorithms. Also the form factor of head mounted displays requires minimal hardware, and low power devices.

Another challenging component for realistic immersive environments in XR systems, is the creation of realistic objects from Photogrammetry and LIDAR data that requires large amounts of storage. New approaches to the generation of 3D scenes using this data are needed.

3 Research program

We will address these problems in the following PhD projects, where we focus on our strengths, collaborate around our shared goals and competencies, and work together on our use of Rendering and Deep Learning techniques.

3.1 PhD-project at Lund University

The PhD project at Lund University will focus on new algorithms for image generation for XR using a combination of Ray Tracing and Deep Learning. We will investigate new sparse Ray Tracing techniques that generate pixels at a much lower resolution, at irregular, important locations in the final image that are used as input to Neural Networks capable of synthesizing the remaining pixels to match very high resolution XR displays. Sparse samples can be generated adaptively and driven by requirements of the Neural Networks, similar to Kuznetsov *et al.* [8] and recent work by Hasselgren *et al.* [5], done by our collaborators at NVIDIA Research Lund.

The recent Ampere GPU Architecture [12] from NVIDIA improves the performance of ray tracing, which is fundamental to the underlying performance of image generation. At the same time it advances the performance of Deep Learning, giving us the best technology to achieve the high frame rates and pixel resolutions required.

To generate pixels from sparsely sampled data we will use Neural Networks. Deep Learning has been used to both upscale and de-noise images generated using physically based techniques such as Path Tracing [2]. We will use similar techniques while adopting recent advances for XR devices such as DeepFovea [7] and Neural Supersampling [17]. Different to these techniques, we will use synthetic datasets for training following a procedural approach, and using open source software as much as possible, in a similar fashion to recent work from our partners at Linköping [16]. We also plan to release the scripts to generate the training data via open source, enabling future researchers to modify and train using the same setup and data. We will investigate methods to improve the computation of network inputs, such as motion vectors, using our Ray Tracing approach.

We will survey the available technology in the XR space, and over the course of the project be ready to work with the newest equipment. We will consider the latest headsets for AR, such as Microsoft's HoloLens, and for VR, Vargo, Value's Index, and Oculus' Quest. We will investigate both the use of off the shelf game engines such as Unreal Engine[1] and Unity[2], as well as using our own Ray Tracing to ensure the optimal user experiences for the target applications. We will target applications in collaboration with PhD project 2, which will combine the neural procedural texturing and industrial applications. To ensure user focused results, we will conduct perceptual user studies with existing technologies.

3.2 PhD-project at Linköping University

The aim for the PhD research project at Linköping University is to develop novel methods to synthesize high-detail in models and textures generated by photogrammetry or from low-resolution data, such as building outlines from Lantmäteriet (the Swedish National Land Survey authority). Photogrammetry, producing 3D models and textures from multi-view photographs, is frequently used to recreate real-world objects and structures in digital form. Objects may be individual artifacts, from decimeter size to a room interior, and even up to an entire city. Often the technique is combined with laser scanning to more accurately determine relative 3D positions of pixels in the images. Despite generating dense 3D meshes and texture images, a 3D model of Norrköping from 2016 is about 70 GB, the quality is quite poor when scrutinized up close (ground level). It is thus imperative to investigate novel techniques and methods to compact the representation of such large models of real-world scenery while significantly increasing the level of detail several magnitudes. Procedural Content Generation (PCG) has been used to create digital content in a wide variety of computer graphics application, traditionally using rules and procedures, and more recently deep learning approaches have been incorporated, as recently surveyed by Liu *et al.* [9]. Our goal is to be able to use less data, both in terms of mesh resolution and amount of texture data, yet achieve higher quality by means of PCG, specifically using procedural textures, combined with a set of deep learning techniques at different levels to produce imagery faithful to the real-world we are aiming to represent.

To this end we will investigate novel approaches using a combination with deep neural features for model selection and parameter optimization using differential material graphs for procedural texture synthesis of surfaces, as recently presented by Shi *et al.* [14]. Our approach will differ as we will investigate a combined use of source image data from photogrammetry, of far lower resolution, for model selection and parameter estimation. In addition, we will investigate the use of additional metadata about the surface area and property information of buildings, data under development at Norrköping and Gotheburg municipalities, for instance: documented roofing material.

Another issue that arises with tiled textures, static and procedural alike, is a visible repetition pattern when the surfaces are viewed from a distance. Thus, we will investigate another novel technique, presented by Frühstück *et al.* [3], that supports creating massive high-detailed images and blend mixtures of tiles. The approach here is, again, to combine it with source images from photogrammetry to direct the variation of the tiling and create a seemingly larger virtual texture and combine it with the procedural texture generation approach above.

We will exploit source images used for photogrammetry (resolution 5-20cm/pixel), captured from low altitude aircrafts as well as close-to-ground drones as input data, together with high-resolution exemplar images (less than 1mm/pixel) and metadata, such as wall type, roofing, materials, and other classifying properties in GIS databases. To this end we will have access to both Norrköping and Gothenburg city data. In addition we will collect material photographs from both cities to use as exemplar high-resolution images, furthermore we can make use Quixel Megascans[3] assets library and other libraries of procedural materials, such as Substance Source[4].

Throughout the project we intend to make source code and scripts available as open source and ready to use plugins for a suitable game engine, such as Unreal Engine by Epic Games.

4 Novelty of the technical approach

The novelty of our approach is the combination of sparse, adaptive Ray Tracing with recently released Ray Tracing hardware. Ray Tracing enables sparse sampling of 3D data, unlike traditional Rasterization based graphics hardware which is required to use uniform grids to sample triangles one at a time. Combining the sparse sampling of Ray Tracing with the ability of trained Neural Networks to generate pixels for higher resolution, pixels in the periphery, and realistic objects, we will be able to generate more pixels than previous methods. One challenge for Path

[1] https://www.unrealengine.com [2] https://unity.com [3] https://quixel.com/megascans/home [4] https://source.substance3d.com

Tracing is coming up with new approaches that can scale older, single threaded approaches, to the massively parallel streaming data processing of modern Ray Tracing hardware. Another novel aspect is the combination of feeding sparse Ray Tracing data into Deep Learning networks for upscaling and foveation. Unlike most previous approaches to training with synthetic data, we will use Physically Based Path Traced data. This will generate more physically realistic imagery, making the XR experience more immersive.

The novel approach we are aiming for is the combination of metadata in conjunction with low to medium resolution image data, originally captured for photogrammetry, to generate high quality and high resolution of textures using procedural textures. By using high quality Photogrammetry and LIDAR datasets to train Neural Networks, we will be able to generate the physically accurate, and high resolution objects needed for interactive exploration using XR. Furthermore we will use procedural texturing and modelling to enable tailoring the synthetic scenes to particular use cases, to ensure the Neural Network upscaling is a good match for its target application.

5 State of the art in an international context

This project will build on recent work at Facebook Reality Labs to use Deep Learning for pixel generation for foveated rendering (DeepFovea), and upscaling (Neural Supersampling). Also recent early work on foveated ray-tracing has shown good results. LU's Graphics group recent work in realistic image generation using Path Tracing, has included one PhD Student's work in collaboration with NVIDIA Research Lund [10, 11], which has lead to further research [1] which has been recently released for developers as RTXDI. This Path Tracing builds the foundations for our sparse sampling Ray Tracing work for this project. Michael spent 2018-19 working at Facebook Reality Labs on Ray Tracing Hardware for XR [13], working in the same team that worked on DeepFovea and Neural Supersampling. Other recent physically based Path Tracing[15] and previous work on acceleration structures for Ray Tracing [4] will impact our research in this project.

We will also build upon the techniques for inverse procedural texture mapping by Hu *et al.* [6] and differential material graphs by Shi *et al.* [14], as well as the creation of seamless tile blending using GAN:s, as presented by Frühstück *et al.* [3], to support creating necessary tile variation over large areas. Microsoft Flight Simulator 2020 showcases that combinations of procedural modeling and texturing in combination with metadata, such as OpenStreetMap, can indeed produce enriched scenery.

6 Relation to and complementarity to other initiatives

At LU there are two projects in Real-Time Ray Tracing, in particular using Path Tracing, which are funded by VR and WASP. Visual Feature Based Data Reduction (B12) explores new representations, and the rendering of them, this project goes further by synthesizing pixels using Neural Networks, and solving the challenges of future XR platforms.

7 Synergy between sites and forms of collaboration

As the project brings together different efforts on two different sites we will establish regular monthly virtual meetings of all partners in a seminar-style. There the partners (Ph.D. students and advisers) will be presenting the current status of their research or papers that are relevant for the project. We will also work with inter-site co-supervision of the PhD students and as a part of this organize extended exchange visits between the sites for the students.

8 Research team

At Lund University two PhDs are actively working with Real-Time Ray Tracing, Pierre Moreau and Gustaf Waldemarson (ARM industrial student), and will be able to offer expertise in the area. Lund University continues to maintain a close collaboration with Tomas Akenine-Möller's group at NVIDIA Research Lund. The Media and Information Technology division (MIT), headed by Anders Ynnerman, at the Department for Science and Technology, Linköping University, hosts some 60 researchers and PhD students within visualization, computer graphics, image processing, computer human interaction and visual learning that may assist in a wide field of related topics. The division holds currently over 80 staff members. Jonas Unger, head of the Computer Graphics and Image Processing group at MIT/ITN is also available for advise on Deep Learning topics. Anders Ynnerman is a Wallenberg Scholar and will support this project as an advisor. Patric Ljung is the head of the Immersive Visualization group at MIT with 16 research engineers and researchers. Norrköping municipality, city planning office, offers access to their photogrammetry data for this research project, both raw data and generated 3D models and textures. We also have an established relation with Gothenburg municipality and their city planning office, which are also offering access to their data.

References

[1] B. Bitterli, C. Wyman, M. Pharr, P. Shirley, A. Lefohn, and W. Jarosz. Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, 39(4), July 2020.

[2] C. R. A. Chaitanya, A. S. Kaplanyan, C. Schied, M. Salvi, A. Lefohn, D. Nowrouzezahrai, and T. Aila. Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder. *ACM Trans. Graph.*, 36(4), July 2017.

[3] A. Frühstück, I. Alhashim, and P. Wonka. Tilegan: Synthesis of large-scale non-homogeneous textures. *ACM Trans. Graph.*, 38(4), July 2019.

[4] P. Ganestam and M. Doggett. SAH guided spatial split partitioning for fast BVH construction. *Computer Graphics Forum (EuroGraphics 2016)*, 35(2):285–293, may 2016.

[5] J. Hasselgren, J. Munkberg, M. Salvi, A. Patney, and A. Lefohn. Neural temporal adaptive sampling and denoising. *Computer Graphics Forum*, 39(2):147–155, 2020.

[6] Y. Hu, J. Dorsey, and H. Rushmeier. A novel framework for inverse procedural texture modeling. *ACM Trans. Graph.*, 38(6), Nov. 2019.

[7] A. S. Kaplanyan, A. Sochenov, T. Leimkühler, M. Okunev, T. Goodall, and G. Rufo. Deepfovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Trans. Graph.*, 38(6), Nov. 2019.

[8] A. Kuznetsov, N. K. Kalantari, and R. Ramamoorthi. Deep adaptive sampling for low sample count rendering. *Computer Graphics Forum*, 37:35–44, 2018.

[9] J. Liu, S. Snodgrass, A. Khalifa, S. Risi, G. N. Yannakakis, and J. Togelius. Deep learning for procedural content generation. *Neural Computing and Applications*, pages 1–19, 2020.

[10] P. Moreau and P. Clarberg. *Importance Sampling of Many Lights on the GPU*, pages 255–283. 2019.

[11] P. Moreau, M. Pharr, and P. Clarberg. Dynamic Many-Light Sampling for Real-Time Ray Tracing. In *High-Performance Graphics - Short Papers*, pages 21–26, 2019.

[12] NVIDIA. *Ampere GA102 GPU Architecture*, 2020.

[13] K. Rajan, S. Hashemi, U. Karpuzcu, M. Doggett, and S. Reda. Dual-precision fixed-point arithmetic for low-power ray-triangle intersections. *Computers & Graphics*, 87:72 – 79, 2020.

[14] L. Shi, B. Li, M. Hašan, K. Sunkavalli, T. Boubekeur, R. Mech, and W. Matusik. Match: Differentiable material graphs for procedural material capture. *ACM Trans. Graph.*, 39(6), Nov. 2020.

[15] G. Waldemarson and M. Doggett. Photon Mapping Superluminal Particles. In *Eurographics 2020 - Short Papers*. The Eurographics Association, 2020.

[16] M. Wrenninge and J. Unger. Synscapes: A photorealistic synthetic dataset for street scene parsing. *CoRR*, abs/1810.08705, 2018.

[17] L. Xiao, S. Nouri, M. Chapman, A. Fix, D. Lanman, and A. Kaplanyan. Neural supersampling for real-time rendering. *ACM Trans. Graph.*, 39(4), July 2020.