Aaryaman Vasishta

linkedin.com/in/adyaman | aaryaman.vasishta@gmail.com | aaryaman.net | +44 07586572385

INTERESTS

- Real-time Path Tracing • FOSS
- Machine Learning • Neural Rendering
- GPU Kernel Optimization

EDUCATION

UNIVERSITY OF TOKYO
MASTERS IN INFORMATION SCIENCE AND TECHNOLOGY
2019-2021 | Tokyo, JP
Computer Graphics Group
Thesis: MCMC using Neural Networks
Advisor: Toshiya Hachisuka

PUNE UNIVERSITY
BE IN COMPUTER ENGINEERING
2012-2016 | Pune, IN
First Class with Distinction
Pune Institute of Computer Technology

LINKS

Github:// **jammm**
BitBucket:// **jammm**

SKILLS

PROGRAMMING
- C • C++ • Python • MATLAB • Lua
- Bash • x86 Assembly • Kotlin

VERSION CONTROL
Git • Perforce

TOOLS
- vim • Visual Studio • Slurm
- OpenGL • HIP • CUDA
- MySQL • Pytorch
- Cassandra • Kubernetes • Docker

OPEN SOURCE

GOOGLE SUMMER OF CODE
Mentor - 2018 WineHQ, 2020 PCL
Student in 2015 and 2016 for WineHQ

MISC. OSS CONTRIBUTIONS
- Mitsuba 2 • PBRTv4 • nouveau
- Chromium • Point Cloud Library
- Appleseed • Zandronum • ScummVM
- SHARK • llama.cpp • llamafile

AWARDS

2023	AMD Executive Spotlight - Stable Diffusion Windows enablement
2023	AMD Executive Spotlight - SHARK Windows ROCm backend enablement
2023	AMD Executive Spotlight - For contributing to AI accelerated demos - RDNA3 launch event.
2023	AMD spotlight - RDNA3 launch event - achieved 2.7x AI performance uplift vs. RDNA2
2020	Japan Student Services Organization (JASSO) scholarship award
2018	Rakuten Tech Division top 1% award with stock bonus

EXPERIENCE

STABILITY AI MACHINE LEARNING ENGINEER | LONDON, UK
March 2024 - Present
- **Stable Fast 3D** - enabled MPS and CUDA backends for the texture baker, worked on tri-tri BVH optimizations.
- Dataset generation and preprocessing for 3D and talking-heads AI models.

ADVANCED MICRO DEVICES RESEARCH ENGINEER | TOKYO, JAPAN
April 2021 - March 2024
- **Influenced multiple future hardware architectures** on neural rendering by developing forward-looking workloads and evaluating them under different hardware constraints.
- **Contributed to future GPU (RDNA4) HW enhancements** after performing analysis of online AI/ML neural rendering workloads such as instant-ngp, Neural Radiance Caching. **HW performance uplift projected up-to 30%.**
- **Awarded executive spotlight for enabling the Windows ROCm backend for nod.ai SHARK.** (later acquired by AMD). Demonstrated the viability of the ROCm backend on Windows as a more flexible alternative to the Vulkan backend.
- **Awarded executive spotlight** for writing fully-fused MLP kernels for current and next-gen RDNA3 GPUs, backing the gen-on-gen AI ops **uplift of 2.7x**.
- **Re-implemented instant-ngp (Instant NeRF) from scratch**, which includes **fully-fused MLP kernels with WMMA optimizations (tensor cores), grid encoding and occupancy grids** for in-house research work with competitive performance for RDNA3 GPUs.
- **Ported tiny-cuda-nn and instant-ngp to HIP**, supporting RDNA3 and MI GPUs, utilizing the WMMA ops of RDNA3 and matrix cores of MI. As part of this effort, modified CUTLASS to use HIP using rocWMMA backend.
- Enabled HIP support for llamafile and contributed towards RDNA3 compatibility for llama.cpp on Windows.
- Contributed to HIP-RT project infrastructure (bitcode linking, kernel baking, compilation performance).
- **Ported PBRTv4 from CUDA to HIP** with assistance from HIP-RT team. Integrated support for interactive mode using HIP-OpenGL interop.
- Ported Radeon ProRender from OpenCL and CUDA to HIP within 48 hours and **improved CPU performance up-to 73%, GPU up-to 25%.**
- Wrote a blog on how to use RDNA3's WMMA instructions with sample code - `https://gpuopen.com/learn/wmma_on_rdna3/` - well received both internally and externally.
- Released Orochi - single-binary runtime linking for both HIP and CUDA. Hardened HIP runtime by writing test cases across HIP and CUDA to fix HIP issues using black-box reverse engineering.
- Investigated UE5 nanite cluster building and Luma real-time NeRF rendering.

RAKUTEN | Software Architect/Lead Engineer | Tokyo, Japan
October 2016 – March 2019
- Awarded for being among the top 1% of the Technology Divison employees.
- **Mentored and onboarded new engineers and interns** from top Canadian universities. Mentored several new engineers and interns from project kickoff and infra setup until deployment in production.
- Designed, developed and deployed large scale, zero-downtime, cloud-native core identity services utilizing **Kubernetes and Cassandra** serving **billions of worldwide Rakuten logins daily**. Overall **savings of ¥12 million annually**.
- RIT: Deep learning for **Soft-segment background removal** of e-commerce images.

UBISOFT | Intern Gameplay Programmer, Pune studio
January 2016 – May 2016
- **Worked on gameplay and engine layers,** fixing issues while remastering **South Park™: The Stick of Truth™** to PS4 and Xbox One.

PROJECTS

PATH TRACER | Physically based renderer used for research
2019 - Present

- **Cross-Platform and Written from scratch** using C++17.
- Current features: Live preview, multi-threaded rendering, Parallel SAH BVH and SBVH. Integrators: Path Tracing with Next-Event-Estimation and MIS, PSS-MLT, PRT using Spherical Harmonics.
- BSDFs supported: Diffuse, Phong, Rough conductor using GGX/Beckmann microfacet model, Dielectric and Metal.

WINEHQ | Open-Source Compatibility layer for running Windows programs on POSIX-compliant OS's
2015 - 2018
- **Re-implemented** Microsoft's Direct3D Retained Mode, **a 3D scene graph API** in C resulting in **improved compatibility** across legacy applications and games. Contributions included in Valve's Proton layer for Steam on Linux.
- Wrote tests against Microsoft's undocumented API while adhering to **black-box reverse engineering** methodology.

PUBLICATIONS

- Daniel Meister, Paritosh Kulkarni, **Aaryaman Vasishta**, and Takahiro Harada. 2024. HIPRT: A Ray Tracing Framework in HIP. Proc. ACM Comput. Graph. Interact. Tech. 7, 3, Article 44 (August 2024), 18 pages. https://doi.org/10.1145/3675378
- Mark Boss, Zixuan Huang, **Aaryaman Vasishta** and Varun Jampani, 2024. SF3D: Stable Fast 3D Mesh Reconstruction with UV-unwrapping and Illumination Disentanglement, arXiv https://arxiv.org/abs/2408.00653