DESIGN OF A PRIVACY-AWARE TEMPORAL SUPERSAMPLING ARCHITECTURE

AKANKSHA DIXIT

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

APRIL 2026

DESIGN OF A PRIVACY-AWARE TEMPORAL SUPERSAMPLING ARCHITECTURE

by

AKANKSHA DIXIT

Department of Electrical Engineering

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

APRIL 2026

Certificate

This is to certify that the thesis titled **DESIGN OF A PRIVACY-AWARE TEMPORAL SUPER-SAMPLING ARCHITECTURE** being submitted by **Ms. AKANKSHA DIXIT** for the award of **Doctor of Philosophy** in Electrical Engineering is a record of bona fide work carried out by her under my guidance and supervision at the Department of Electrical Engineering, Indian Institute of Technology Delhi. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Smruti R. Sarangi
Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi- 110016

Acknowledgements

I would like to start with expressing my deepest gratitude to **Prof. Smruti R. Sarangi** for his unwavering guidance and support throughout my PhD. His relentless pursuit of excellence inspired me to push my limits and grow both academically and personally. A crucial lesson I learned from him is the importance of maintaining a big-picture perspective and addressing problems with a long-term vision. This has transformed my approach to problem-solving, enabling me to tackle complex issues more effectively.

I would also like to thank my Student Research Committee: **Prof. Sumantra Dutta Roy**, **Prof. Seshan Srirangarajan**, and **Prof. Rohan Paul**, for their insightful comments, constructive feedback, and continuous guidance during the course of my research.

I am grateful to my friends at IIT Delhi for many wonderful conversations, both technical and non-technical. I especially appreciate my seniors and friends, Dr. Priyanka Singla and Dr. Nivedita Shrivastava, for their advice on planning research and tackling challenges, as well as their encouragement during difficult times. I want to extend a special thanks to my labmates, who have become dear friends: Shruti, Ani, and Rohith, for always being there and making this journey enjoyable and memorable.

I am deeply grateful to my parents for their unconditional love, constant support, and unwavering belief in me. Their encouragement allowed me to devote myself fully to this research. They are the pillars of my life and have always reminded me that with hard work, anything is possible and that I should focus on the effort rather than the outcome. Above all, I thank GOD for all and everything.

I also acknowledge the use of OpenAI's ChatGPT to assist with language refinement in this thesis.

Akanksha Dixit

Abstract

Graphics applications have become an integral part of diverse domains, ranging from entertainment and education to healthcare and scientific visualization. These applications take various forms. They can be immersive such as virtual reality (VR) and augmented reality (AR), or non-immersive, such as traditional desktop-based applications. They can be deployed across multiple platforms such as desktops, mobiles and headsets. Each form and platform presents unique performance constraints, yet a universal expectation remains: delivering a high Quality of Experience (QoE) to the users. Meeting this expectation requires addressing multiple challenges: ensuring low latency and high frame rendering rates, achieving high visual quality, maintaining realism, and safeguarding user security and privacy. This thesis takes a holistic approach to meeting these requirements while operating within the constraints of the underlying hardware resources.

We begin with the problem of high latency in VR systems, specifically the motion-to-photon delay (MPD), which is the time between a user's head movement and the corresponding update on the display. In VR, even slight delays can disrupt the user's sense of presence, making the MPD more critical. A common MPD reduction technique is Asynchronous Time Warping (ATW), which reprojects the already rendered frame based on the most recent head position. This helps display the frame corresponding to the latest possible head position, reducing the MPD. However, the challenge is to invoke ATW at the right instance within the refresh cycle; otherwise, it will increase the MPD. Fixing a time for invoking ATW does not work because ATW shares resources with the rendering pipeline, leading to unpredictable latency due to preemption and context switches. Hence, we propose, *PredATW*, a predictor that monitors the status of hardware resources and the rendering process to accurately estimate ATW latency, thereby reducing MPD and minimizing missed deadlines. Our predictor has an error of only 0.22 ms across several popular VR applications. Moreover, as compared to the baseline architecture, we reduce deadline misses by 80.6%.

While *PredATW* reduces the MPD by correctly invoking ATW, it is important to note that ATW has its own limitations. Specifically, it only corrects for head rotation, not translation, and primarily enhances perceived responsiveness by utilizing the most recent position data. However, it does not increase the frame rate. This makes ATW effective at relatively modest refresh rates spanning from 60 to 90 Hz, which are typical for VR headsets, but it falls short for modern monitors with high refresh rates (240 to 360 Hz) that require higher frame rates. To increase the frame rate, techniques such as *frame generation* or *temporal supersampling* have been proposed, which predict intermediate frames from already rendered frames.

Existing temporal supersampling methods are *interpolation* and *extrapolation*. Interpolation, which uses both past and future frames, offers higher quality but is best for doubling frame rates, as further upsampling adds latency and complexity. Extrapolation relies only on past frames, enabling continuous frame generation with lower latency, making it more suitable for real-time applications. In this work, we target a $4\times$ frame rate increase (from 60 Hz to 240 Hz), making extrapolation the preferred choice. However, no single existing extrapolation algorithm performs consistently well across all scene types. Each one is tailored to specific scenarios and exhibits strengths and weaknesses depending on the context. Hence, we propose *X4Rate*, a frame generation framework that combines multiple existing methods to deliver results far superior to any individual algorithm by selecting the best algorithm for the current scene. This adaptive selection ensures high visual quality consistently because when one method underperforms, others compensate. *X4Rate* achieves $4\times$ the rendering rate while delivering high-quality frames, outperforming the nearest competitor, ExtraNet, with a 22.13% improvement in the PSNR.

Building on this, we address the inherent limitation of relying solely on existing extrapolation methods by designing our own. We propose a novel lightweight frame extrapolation algorithm that exploits the perceptual sensitivity to segment each frame into foreground and background regions and employs a novel neural network to generate the final extrapolated frame. Additionally, a wavelet transform (WT)-based filter pruning technique is applied to compress the network, significantly reducing the runtime of the extrapolation process. Our results demonstrate that *PatchEX* achieves a 61.32% and 49.21% improvement in PSNR over the latest extrapolation methods ExtraNet and ExtraSS, respectively, while being $3\times$ and $2.6\times$ faster (resp.).

Finally, we address the data security and privacy challenges in AR systems. By blending the real and virtual worlds, AR devices can inadvertently expose sensitive visual data—ranging from bystanders' identities to confidential physical documents. Existing privacy-preserving methods are often too computationally heavy for real-time execution on resource-constrained AR headsets. To solve this, we introduce *ARGuard*, a lightweight, real-time framework that integrates

a saliency-aware object detection network with a novel irreversible obfuscation algorithm. *AR-Guard* operates at 60 fps on Microsoft HoloLens 2, detecting and protecting sensitive regions with 12.6% and 15.8% higher accuracy in detection and tracking, respectively, and achieving a 60.23% reduction in privacy leakage compared to the nearest competing method, *BystandAR*.

To summarize, this thesis presents a unified framework for addressing both performance and privacy in graphics applications. Through innovations in latency reduction, high-quality extrapolation, and real-time privacy preservation, our work advances the state-of-the-art in delivering secure, smooth, and immersive experiences on consumer-grade hardware.

संक्षेप

मनोरंजन और शिक्षा से लेकर स्वास्थ्य सेवा और वैज्ञानिक विज़ुअलाइज़ेशन तक, ग्राफ़िक्स अनुप्रयोग विविध क्षेत्रों का एक अभिन्न अंग बन गए हैं। ये अनुप्रयोग विभिन्न रूप धारण करते हैं, जिन्हें इमर्सिव के रूप में वर्गीकृत किया जाता है, जैसे वर्चुअल रियलिटी (वी आर), ऑगमेंटेड रियलिटी (AR), या गैर-इमर्सिव, जैसे पारंपरिक डेस्कटॉप-आधारित अनुप्रयोग। ये अनुप्रयोग डेस्कटॉप, मोबाइल और हेडसेट सहित कई प्लेटफ़ॉर्म पर तैनात किए जाते हैं। प्रत्येक रूप और प्लेटफ़ॉर्म अद्वितीय प्रदर्शन बाधाएँ प्रस्तुत करता है, फिर भी एक सार्वभौमिक अपेक्षा बनी रहती है: उपयोगकर्ताओं को उच्च अनुभव गुणवत्ता (QoE) प्रदान करना। इस अपेक्षा को पूरा करने के लिए कई चुनौतियों का समाधान करना आवश्यक है: कम विलंबता और उच्च फ़्रेम रेंडरिंग दर सुनिश्चित करना, उच्च दृश्य गुणवत्ता प्राप्त करना, यथार्थवाद बनाए रखना और उपयोगकर्ता सुरक्षा और गोपनीयता की रक्षा करना। यह शोध प्रबंध अंतर्निहित हार्डवेयर संसाधनों की सीमाओं के भीतर काम करते हुए इन आवश्यकताओं को पूरा करने के लिए एक समग्र दृष्टिकोण अपनाता है।

हम वी आर सिस्टम में उच्च विलंबता की समस्या से शुरुआत करते हैं, विशेष रूप से मोशन-टू-फ़ोटॉन विलंब (एम-पी-डी), जो उपयोगकर्ता के सिर की गति और डिस्प्ले पर संबंधित अपडेट के बीच का समय है। वी आर में, थोड़ी सी भी देरी उपयोगकर्ता की उपस्थिति की भावना को बाधित कर सकती है, जिससे एम-पी-डी और भी महत्वपूर्ण हो जाता है। एम-पी-डी कम करने की एक सामान्य तकनीक एसिंक्रोनस टाइम वॉर्पिंग (ए-टी-डब्ल्यू) है, जो पहले से रेंडर किए गए फ़्रेम को सबसे हालिया हेड पोजीशन के आधार पर रीप्रोजेक्ट करती है। इससे एम-पी-डी को कम करते हुए, नवीनतम संभावित हेड पोजीशन के अनुरूप फ़्रेम को प्रदर्शित करने में मदद मिलती है। हालाँकि, चुनौती रिफ़्रेश चक्र के भीतर सही समय पर ए-टी-डब्ल्यू को ट्रिगर करने में है;अन्यथा, यह अभी भी एम-पी-डी को बढ़ाता है और सबसे खराब स्थिति में रिफ़्रेश की समय सीमा चूक सकता है। निश्चित समय निर्धारण विश्वसनीय रूप से काम नहीं करता क्योंकि ए-टी-डब्ल्यू रेंडरिंग पाइपलाइन के साथ संसाधन साझा करता है, जिससे प्रीएम्पशन और संदर्भ स्विच के कारण अप्रत्याशित देरी होती है। इसलिए, हम Predए-टी-डब्ल्यू, एक ऐसा प्रेडिक्टर प्रस्तावित करते हैं जो हार्डवेयर संसाधन की स्थिति और रेंडरिंग कार्यभार की निगरानी करके ए-टी-डब्ल्यू विलंबता का सटीक अनुमान लगाता है, जिससे एम-पी-डी कम होता है और छूटी हुई समय-सीमाएँ न्यूनतम होती हैं। हमारे प्रेडिक्टर में कई लोकप्रिय वी आर अनुप्रयोगों में केवल 0.22 ms की त्रुटि है। साथ ही, बेसलाइन आर्किटेक्चर की तुलना में, हम समय-सीमा चूक को 80.6% तक कम करते हैं।

जहाँ वी आर हेडसेट आमतौर पर अपेक्षाकृत कम रिफ़्रेश दरों पर काम करते हैं, वहीं आधुनिक मॉनिटर अब 360 हर्ट्ज़ तक पहुँच जाते हैं। हालाँकि ये डिस्प्ले ज़्यादा स्मूथ मोशन और कम आर्टिफैक्ट्स जैसे कि जडर और मोशन ब्लर का वादा करते हैं, लेकिन अगर जी-पी-यू रिफ़्रेश दर के बराबर तेज़ी से फ़्रेम रेंडर नहीं कर पाता है, तो इनके फ़ायदे सीमित हो जाते हैं और यहाँ तक कि नवीनतम जी-पी-यू भी इतनी तेज़ रेंडरिंग गति को लगातार बनाए रखने में संघर्ष करते हैं। जब रेंडरिंग और रिफ़्रेश दरें बेमेल होती हैं, तो उपयोगकर्ताओं को स्क्रीन फटना और रुक-रुक कर चलने का सामना करना पड़ सकता है। इस समस्या के समाधान के लिए, फ़्रेम जनरेशन या टेम्पोरल सुपरसैंपलिंग जैसी तकनीकें प्रस्तावित की गई हैं, जो पहले से रेंडर किए गए फ़्रेमों से मध्यवर्ती फ़्रेमों का अनुमान लगाती हैं, जिससे प्रत्येक फ़्रेम के लिए पूरी रेंडरिंग पाइपलाइन को निष्पादित किए बिना अनुमानित फ़्रेम दर बढ़ जाती है। मौजूदा टेम्पोरल

सुपरसैंपलिंग विधियाँ इंटरपोलेशन और एक्सट्रपोलेशन हैं। इंटरपोलेशन, जो पिछले और भविष्य दोनों फ़्रेमों का उपयोग करता है, उच्च गुणवत्ता प्रदान करता है लेकिन फ़्रेम दरों को दोगुना करने के लिए सबसे अच्छा है, क्योंकि आगे अपसैंपलिंग से विलंबता और जटिलता बढ़ जाती है। एक्सट्रपलेशन केवल पिछले फ़्रेमों पर निर्भर करता है, जो कम विलंबता के साथ निरंतर फ़्रेम निर्माण को सक्षम बनाता है, जिससे यह वास्तविक समय के अनुप्रयोगों के लिए अधिक उपयुक्त हो जाता है। इस कार्य में, हमारा लक्ष्य 4x फ़्रेम दर वृद्धि (60 हर्ट्ज़ से 240 हर्ट्ज़ तक) है, जिससे एक्सट्रपलेशन पसंदीदा विकल्प बन जाता है। हालाँकि, कोई भी मौजूदा एक्सट्रपलेशन एल्गोरिथम सभी प्रकार के दृश्यों में लगातार अच्छा प्रदर्शन नहीं करता है। प्रत्येक एल्गोरिथम विशिष्ट परिदृश्यों के लिए तैयार किया गया है और संदर्भ के आधार पर अपनी शक्तियाँ और कमज़ोरियाँ प्रदर्शित करता है। इसलिए, हम X4Rate का प्रस्ताव करते हैं, जो एक फ़्रेम निर्माण ढाँचा है जो वर्तमान दृश्य के लिए सर्वोत्तम एल्गोरिथम का चयन करके किसी भी व्यक्तिगत एल्गोरिथम से बेहतर परिणाम देने के लिए कई मौजूदा विधियों को जोड़ता है। यह मूल रूप से एक निर्णय पूर्वसूचक (मेटा-लर्नर) है जो रनटाइम पर गतिशील रूप से सर्वश्रेष्ठ प्रदर्शन करने वाले फ़्रेम एक्सट्रपलेशन एल्गोरिथम का चयन करता है। यह अनुकूली चयन लगातार उच्च दृश्य गुणवत्ता सुनिश्चित करता है क्योंकि जब एक विधि कम प्रदर्शन करती है, तो अन्य उसकी भरपाई कर देते हैं। X4Rate उच्च-गुणवत्ता वाले फ़्रेम प्रदान करते हुए 4 गुना रेंडरिंग दर प्राप्त करता है, जो निकटतम प्रतिद्वंद्वी, एक्स्ट्रानेट से बेहतर प्रदर्शन करता है, और पी एस एन आर में 22.13% सुधार करता है।

इसी आधार पर, हम अपनी स्वयं की एक्सट्रपलेशन विधियाँ डिज़ाइन करके केवल मौजूदा एक्सट्रपलेशन विधियों पर निर्भर रहने की अंतर्निहित सीमा का समाधान करते हैं। हम एक नवीन हल्के फ़्रेम एक्सट्रपलेशन एल्गोरिथम प्रस्तावित करते हैं जो प्रत्येक फ़्रेम को अग्रभूमि और पृष्ठभूमि क्षेत्रों में विभाजित करने के लिए अवधारणात्मक संवेदनशीलता का उपयोग करता है और अंतिम एक्सट्रपलेशन फ़्रेम उत्पन्न करने के लिए एक नवीन तंत्रिका नेटवर्क का उपयोग करता है। इसके अतिरिक्त, नेटवर्क को संपीड़ित करने के लिए एक वेवलेट ट्रांसफ़ॉर्म-आधारित फ़िल्टर प्रूनिंग तकनीक लागू की जाती है, जिससे एक्सट्रपलेशन प्रक्रिया का रनटाइम काफी कम हो जाता है। हमारे परिणाम दर्शाते हैं कि PatchEX नवीनतम एक्सट्रपलेशन विधियों एक्स्ट्रानेट और एक्स्ट्राएसएस की तुलना में क्रमशः 61.32% और 49.21% पी एस एन आर में सुधार प्राप्त करता है, जबकि क्रमशः 3 गुना और 2.6 गुना तेज़ है।

अंत में, हम AR प्रणालियों में डेटा सुरक्षा और गोपनीयता चुनौतियों का समाधान करते हैं। वास्तविक और आभासी दुनिया को मिलाकर, AR उपकरण अनजाने में संवेदनशील दृश्य डेटा को उजागर कर सकते हैं—जिसमें दर्शकों की पहचान से लेकर गोपनीय भौतिक दस्तावेज़ तक शामिल हैं। मौजूदा गोपनीयता-संरक्षण विधियाँ अक्सर संसाधन-सीमित AR हेडसेट पर वास्तविक समय में निष्पादन के लिए कम्प्यूटेशनल रूप से बहुत भारी होती हैं। इसे हल करने के लिए, हम ARGuard प्रस्तुत करते हैं, जो एक हल्का, वास्तविक समय ढाँचा है जो एक प्रमुखता-जागरूक वस्तु पहचान नेटवर्क को एक नए अपरिवर्तनीय अस्पष्टीकरण एल्गोरिथम के साथ एकीकृत करता है। ARGuard माइक्रोसॉफ़्ट होलोलेन्स 2 पर 60 fps पर संचालित होता है, संवेदनशील क्षेत्रों का पता लगाता और उनकी सुरक्षा करता है| क्रमशः 12.6% और 15.8% अधिक सटीकता के साथ, पहचान और ट्रैकिंग में, और सबसे अच्छी मौजूदा विधि, BystandAR की तुलना में गोपनीयता रिसाव में 60.23% की कमी प्राप्त करता है।

संक्षेप में, यह शोध-प्रबंध ग्राफ़िक्स अनुप्रयोगों में प्रदर्शन और गोपनीयता दोनों को संबोधित करने वाला एक एकीकृत ढाँचा प्रस्तुत करता है। विलंबता में कमी, उच्च-गुणवत्ता वाले एक्सट्रपलेशन, और

वास्तविक समय गोपनीयता संरक्षण में नवाचारों के माध्यम से, हमारा कार्य उपभोक्ता-स्तरीय हार्डवेयर पर सुरक्षित, सुचारू और इमर्सिव अनुभव प्रदान करने में अत्याधुनिक तकनीक को आगे बढ़ाता है।

Contents

List of Figures

List of Tables