Ultra Low Latency Augmented Reality Application Leveraging Edge Computing

Mustafa Bin Amir, Azmeer Faisal, Muhammad Nameer Anjum

Introduction:

Augmented Reality (AR) involves the integration of digital information into the physical environment. AR enhances the user's environment with virtual elements using cameras, sensors, and other devices. AR can be experienced on smartphones, tablets, smart glasses, and head-mounted displays. To provide a smooth AR experience, several needs must be met, including high-quality screens for projecting virtual objects and precise user tracking, high capacity, low latency computing, low latency communication, reliable communication and scalability. These requirements continue to be a key barrier in AR development, significantly harming user experience and inhibiting mainstream adoption. To solve these issues, researchers and developers are looking at novel ways to minimize latency and improve the AR experience, such as employing edge computing and improving code performance to shorten data transmission times.

Traditional cloud computing involved sending the user device-generated data for processing to a large centralized data center. With the amount of data growing at an exponential rate, it has overwhelmed the traditional data center infrastructure leading to problems such as latency issues and network congestion. Edge computing evolved as a response to cloud computing's shortcomings. It puts computation at a closer physical proximity to the source of data, allowing IoT devices and other applications to do real-time processing and decision-making. It does this by dispersing nodes called cloudlets on the edges of the Internet to allow end user devices to offload computation .Edge computing supplements cloud computing, and the decision relies on the individual needs of each application.

By leveraging edge computing one could overcome key challenges in AR development. It provides an efficient and scalable approach to provide low-latency computing to AR systems.

Related Work:

The Emergence of Edge Computing examines the emerging trend of edge processing, which moves computing closer to the network's edge in order to enhance latency, security, and dependability for new use cases.

Reducing Latency in a Collaborative Augmented Reality Service demonstrates the use of marker-based tracking to develop an 'One-for-All-Shared-Experience' (OFALL-SE) dynamic object localization service, which reduces the communication and computation latency in comparison to the conventional Local Anchor Transfer (LAT) method.

Towards Low-Latency and Ultra-Reliable Virtual Reality addresses the limitations posed by latency, reliability, and capacity in the advancement of virtual reality technology. It highlights how mobile edge computing and caching can effectively mitigate these constraints, thereby enhancing the overall VR experience.

Energy-Efficient Task Offloading and Resource Allocation via Deep Reinforcement Learning for Augmented Reality in Mobile Edge Networks presents a deep reinforcement learning approach to optimize energy efficiency in AR on mobile devices by allocating resources in edge networks.

Efficient Multi-player Computation Offloading for VR Edge-Cloud Computing System presents a network system that uses intelligent core network technology for users with limited computing resources.

Proposed Approach:

Our primary focus is to minimize latency to reduce the response time between our application and the Edge sites. This could be achieved by using such open source libraries and APIs which upon retrieving current state data, also hints the application of the future data state. Furthermore open-source “in-memory-store” will be implemented to cache the data within edge-sites.

The application will be designed to dynamically connect to the nearest during user mobility, and the edge data stores will be capable of interacting and migrating state to reduce latency. To reduce compute and communication latencies, both the application and the edge data storage will be improved.

Tools to be used:

Application: Using libraries and APIs through which AR apps are designed. Moreover to control the latency part within the app, open-source algorithms could be implemented, such as ARcore, ARtoolkit, ARkit, MRTK (Mixed Reality Tool-kit by Microsoft) ,Vuforia etc.

Persistent caching: This will be implemented using already existing open-source “in-memory-store” systems such as Redis, Memcache or Apache Ignite.

Development Methodology:

We will start by looking at the tools and languages needed to create such apps. Following that, prototyping will commence. After developing a working prototype, we will deploy the functions to cloud services. Once we have confirmed that the app works on standard cloud services, we will shift our focus to distributed systems such as Redis and Memcache to design persistent caching of data on cloud so that the app can communicate with the edge-site or cloudlets. In the end, testing and benchmarking will begin to evaluate the app.

Timeline and Division of Work

Instead of tasks, each task's subtask will be shared among group members, resulting in each member working on every task and the project as a whole.

Task Timeline	Start Dates (Tentative)
Literature Review	8 Feb 2023
Application Idea	10 Feb 2023
Feasible Solution	15 Feb 2023
Start Building Prototype App	23 Feb 2023
Mid Report	-
Functions deployment on Cloud	31 March 2023
Task/State migration to Cloudlets	3 April 2023

Testing	10 April 2023
Benchmarking	10 April 2023
Evaluation	21 April 2023
Final Report	28 April 2023

GitHub Repository Link

https://github.com/mustafakhokhar/CS-678-ResearchProj

References:

1. Satyanarayanan, M. (2017). The Emergence of Edge Computing. *IEEE Computer*, *50*(1), 30–39. https://doi.org/10.1109/mc.2017.9
2. He, W., Swift, B., Gardner, H., Xi, M., & Adcock, M. (2019). Reducing Latency in a Collaborative Augmented Reality Service. *Virtual Reality Continuum and Its Applications in Industry*. https://doi.org/10.1145/3359997.3365699
3. Mohammed S. Elbamby, Cristina Perfecto, Mehdi Bennis, & Klaus Doppler (2018). Towards Low-Latency and Ultra-Reliable Virtual Reality
4. Alshahrani, A., Elgendy, I. A., Muthanna, A., AlGhamdi, A. S., & Alshamrani, A. (2020). Efficient Multi-Player Computation Offloading for VR Edge-Cloud Computing Systems. *Applied Sciences*, *10*(16), 5515. https://doi.org/10.3390/app10165515
5. Chen, X., & Liu, G. (2021). Energy-Efficient Task Offloading and Resource Allocation via Deep Reinforcement Learning for Augmented Reality in Mobile Edge Networks. *IEEE Internet of Things Journal*, *8*(13), 10843–10856. https://doi.org/10.1109/jiot.2021.3050804

Group Number: 08

Muhammad Nameer Anjum	24100270
Mustafa Bin Amir	24100084
Azmeer Faisal	24100164