World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:01, 2023

Protocol for Dynamic Load Distributed Low Latency Web-Based Augmented Reality and Virtual Reality

Authors : Rohit T. P., Sahil Athrij, Sasi Gopalan

Abstract : Currently, the content entertainment industry is dominated by mobile devices. As the trends slowly shift towards Augmented/Virtual Reality applications the computational demands on these devices are increasing exponentially and we are already reaching the limits of hardware optimizations. This paper proposes a software solution to this problem. By leveraging the capabilities of cloud computing we can offload the work from mobile devices to dedicated rendering servers that are way more powerful. But this introduces the problem of latency. This paper introduces a protocol that can achieve high-performance low latency Augmented/Virtual Reality experience. There are two parts to the protocol, 1) In-flight compression The main cause of latency in the system is the time required to transmit the camera frame from client to server. The round trip time is directly proportional to the amount of data transmitted. This can therefore be reduced by compressing the frames before sending. Using some standard compression algorithms like JPEG can result in minor size reduction only. Since the images to be compressed are consecutive camera frames there won't be a lot of changes between two consecutive images. So inter-frame compression is preferred. Inter-frame compression can be implemented efficiently using WebGL but the implementation of WebGL limits the precision of floating point numbers to 16bit in most devices. This can introduce noise to the image due to rounding errors, which will add up eventually. This can be solved using an improved interframe compression algorithm. The algorithm detects changes between frames and reuses unchanged pixels from the previous frame. This eliminates the need for floating point subtraction thereby cutting down on noise. The change detection is also improved drastically by taking the weighted average difference of pixels instead of the absolute difference. The kernel weights for this comparison can be fine-tuned to match the type of image to be compressed. 2) Dynamic Load distribution Conventional cloud computing architectures work by offloading as much work as possible to the servers, but this approach can cause a hit on bandwidth and server costs. The most optimal solution is obtained when the device utilizes 100% of its resources and the rest is done by the server. The protocol balances the load between the server and the client by doing a fraction of the computing on the device depending on the power of the device and network conditions. The protocol will be responsible for dynamically partitioning the tasks. Special flags will be used to communicate the workload fraction between the client and the server and will be updated in a constant interval of time (or frames). The whole of the protocol is designed so that it can be client agnostic. Flags are available to the client for resetting the frame, indicating latency, switching mode, etc. The server can react to client-side changes on the fly and adapt accordingly by switching to different pipelines. The server is designed to effectively spread the load and thereby scale horizontally. This is achieved by isolating client connections into different processes.

Keywords : 2D kernelling, augmented reality, cloud computing, dynamic load distribution, immersive experience, mobile computing, motion tracking, protocols, real-time systems, web-based augmented reality application

Conference Title : ICCCITDN 2023 : International Conference on Computer Communications, Information Technology and Data Networking

Conference Location : Singapore, Singapore

Conference Dates : January 09-10, 2023