QoE studies on Interactive 3D Tele-Immersion

Kevin Desai, Suraj Raghuraman, Rong Jin, Balakrishnan Prabhakaran
Department of Computer Science
The University of Texas at Dallas
Richardson, TX, USA
{Kevin.Desai, suraj, Rong.Jin, bprabhakaran}@utdallas.edu

***Abstract*—Users' Quality of Experience (QoE) in Interactive 3D Tele-Immersion (i3DTI) systems is influenced by several factors such as the quality of the "live" 3D avatars of the users, network latency, rendering methodology (head mounted display or regular TV type of display), etc. Hence, it becomes important to answer the question: "Is Visual Quality (VQ) the only factor to be considered or do better immersion and faster interactions matter, in having good QoE?" To answer this question, in this paper, a highly optimized state-of-the-art i3DTI framework implementation is introduced along with a soccer-inspired penalty shootout game. This game allows users to experience various situations, view different angles, perceive delays, track virtual ball motion, and play naturally using their entire bodies. A head mounted display device - Oculus Rift allows the users to get completely immersed and perform better in the penalty shootout game compared to watching themselves play on a 3D TV. This scenario is obtained in a controlled lab setting with ultra-high-speed network that has ultra-low latency, high VQ, fast and realistic interactions. Such ideal conditions are not typically available in a wide area Internet. Hence, for faster interaction scenario, we used lesser RGB-D cameras for 3D reconstruction, thereby reducing the model quality significantly. The high responsiveness of the game masked the user's perception of quality, and resulted in them not noticing the lower VQ of the reconstructed 3D models. Based on the results from the user study that focused on immersion, interaction and VQ aspects; users felt the game to be visually appealing, intuitive, engaging, and highly entertaining in all of the scenarios.**

***Index Terms*—3D Tele-Immersion, Interaction, Quality of Experience, Virtual/Augmented Reality, Gaming**

I. Introduction

3D Tele-Immersion (3DTI) systems enable collaborative augmented virtuality, by allowing geographically distributed users to be captured by multiple RGB-D cameras and rendered inside an immersive virtual world. Interactive 3DTI (i3DTI) system, presented in the paper, allows virtual interactions within the virtual environment on top of showing the 3D model, making the application more appealing and immersive. Each user in an i3DTI system is captured using multiple RGB-D cameras to generate a good quality 3D mesh of the user in every frame. These meshes are then transmitted across network and rendered at different sites. This allows remote users to see, communicate and interact with each other. Even for a single camera, large amount of data (approximately 6MB) is captured, processed, and transmitted every frame, resulting in low frame rates and a perceivable lag even on currently available high speed networks. So, even the current state-of-the-art 3D Tele-Immersion (3DTI) applicatons for gaming [1]–[3] and training [4], [5] are restricted to a single RGB-D camera, thereby limiting the possible in-game experience in terms of virtual camera and user interactions. Capture using multiple cameras is vital to support user customizable viewing options in i3DTI applications. This compounds the latency of the application, prior research [1] recommends a maximum latency of 120ms for ensuring user engagement. To achieve such low latency, compromises in visual quality need to be made. Mesh simplification can reduce the mesh size, but is time consuming and unsuitable for real-time applications. The fastest compression/decompression methods require more than 100ms per mesh, increasing latency significantly.

A. Need for a new study

Quality of Experience (QoE) in i3DTI systems is affected by a lot of factors such as the Visual Quality (VQ) of the live captured 3D models of the users, quality of the virtual scene, network latency and delay, rendering device and view - head mounted display, TV, first person, third person, accurate interactions, better immersion in the virtual environment, etc. Prior studies [6] have focused on the quality of the rendering in 3DTI, or the effects of latency on user engagement in a 3DTI game [1]. Current i3DTI systems are capable of providing high VQ in a controlled lab setup and at the same have low latency. Still, the in-game experience or the overall QoE of the system is not really appreciative. Hence, it is really important to understand what other factors influence the overall QoE, apart from VQ. In this paper, we try to answer the question - Is QoE improved by having better immersion in the virtual environment or by having fast and realistic interactions with the virtual objects?

B. Proposed Approach

Our i3DTI framework uses GPU acceleration to speed up processing, and distributes the work load across multiple machines to reduce processing times drastically, and allow a 4 camera system to respond in under 50ms. Natural full body user interactions are realized by using just the skeleton of the user. A virtual camera view based rendering approach is used to select a small subset of cameras that need to be processed for rendering. Although this reduces the delay significantly, it is still not sufficient in situations where there are many cameras with overlapping views. By determining if users playing an engaging game require, or even notice, the change in quality of their 3D reconstructed models, lower

978-1-5386-2937-6/17 $31.00 © 2017 IEEE
DOI 10.1109/ISM.2017.27

Fig. 1: Third person view of the striker kicking the ball towards the goal in the penalty shootout game.

resolution meshes can be generated to reduce system latency significantly.

A two player penalty shootout game, Figure 1, based on the i3DTI framework is developed to study the users' QoE requirements. The impact of immersion on the users' QoE is studied by considering two different versions of the game. The first version shows the penalty game on a 3D TV, in third person, and the second version shows it on a head mounted display - Oculus Rift, in first person. In both the versions, the VQ is kept high and the system latency is kept low to achieve high realism in the physical interactions. Users rated the gaming aspects and the VQ of the other person collaborating in the game. In turn, we try to understand how the gaming experience is modified by getting better immersed into the environment.

As mentioned before, in an i3DTI system involving multiple cameras and realistic network conditions, achieving both high VQ and low latency is very difficult. Compromises have to be made either in the VQ or in the interaction quality. Another study was performed to understand the impact of VQ and interactions on the overall QoE of the game. Two different versions of the game are created, one having high VQ and the other having high interaction quality. To eliminate the bias due to the other player, a single player version of the game is used for this. Users rated their experience of the game, in terms of the noticeable VQ, realism in the interactions as well as other in-game experiences. From the user study we try to understand the users' preferences in having high interaction realism to obtain a better gaming experience.

The subjective evaluations of the users, along with the system performance and in-game player performance scores, provide following interesting insights while trying to answer the original question about QoE.

- The users' overall QoE in an i3DTI system is not just limited to the VQ of the 3D model.
- Immersing the user better in the virtual environment using head mounted display improves the overall gameplay experience.
- QoE is improved if users' interactions with the objects in the virtual environment are fast and realistic.

II. Related Work

Virtual Reality (VR) applications have been used to study or improve player performance in ball games. The major obstacle for wide scale adoption of such applications is the latency associated with the systems. A comprehensive survey of these efforts can be found in [7].

Approaches for real time 3D reconstruction using multiple cameras in 3DTI have been studied extensively and beyond the scope of this paper. A survey of recent 3DTI related reconstruction approaches and applications is given in [8]. The first 3DTI gaming application "I'm a Jedi" was presented by [1]. This game allowed interaction using a light saber between remote players. A single stereo camera was used per player, and no virtual world objects or physics were used in the game. An immersive tennis game, using body sensors to control the racket, was shown by [2]. A Kinect camera was used to capture the player's back, and a point cloud was rendered using opengl. Basic collision detection was performed to allow the players to hit the ball. The baseball game developed by [3] consisted of two Kinects, each capturing a different player. Virtual objects, like a baseball bat and ball, were placed in the players' hands, based on the joint information provided by the Kinect. Players used gestures to throw the ball and due to limited cameras, only a fixed view of the scene was provided. All the above games use a single camera per site to capture the player, reducing the processing time and overall complexity. The players are rendered inside the scene, but their interaction is restricted to certain joints, or are dependent on external devices like sensors in [2], or the light saber in [1]. Our game allows multiple angle rendering and whole body interaction.

The effect of VQ of the mesh on user experience was studied by [6]. The study focused on how much deterioration is inconceivable by the user, allowing for lower quality meshes to be transmitted without any change in user perception. Rather than letting the user focus only on the rendered information, we engage the users in an interactive and immersive activity and study their response on various questions related to their overall experience, including VQ.

III. i3DTI Framework

An i3DTI system allows geographically distributed users to be present in a virtual collaborative environment and interact with each other and virtual objects using their body. To achieve 3D presence, the user is captured from all directions using multiple calibrated RGB-D cameras simultaneously, to generate a 3D mesh that can be virtually rendered. The i3DTI framework is designed to address the major needs of any i3DTI system, and allow rapid i3DTI application development. The i3DTI framework handles not only the capture, reconstruction, transmission, and rendering of the user, but also enables the natural full body user interaction with the system. For each of

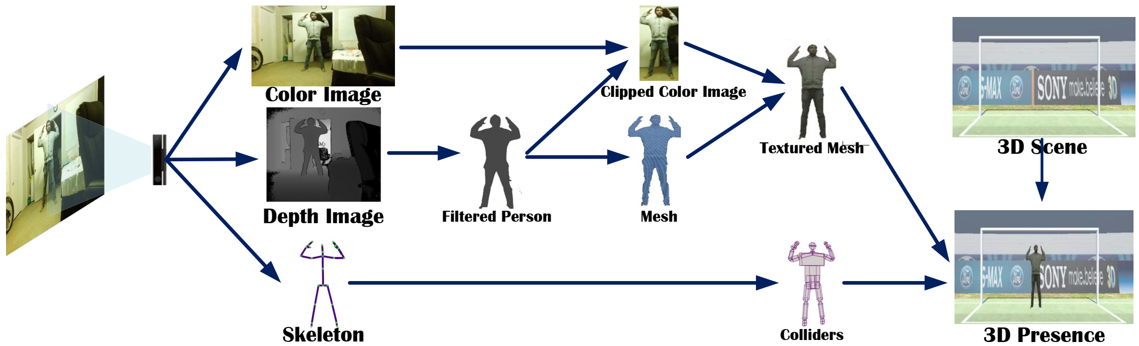


Fig. 2: Reconstruction pipeline for a 3D immersive application using MS Kinect V2 for capture, and Unity3D for rendering.

the cameras, the framework uses a rendering pipeline and an interaction pipeline, as shown in Figure 2.

A. *Rendering*

The depth image returned by the RGB-D cameras is noisy, and the characteristics of the noise vary from camera to camera. Depending on the noise, a median/bilateral or similar filter is applied on the depth image, to improve image quality. The user can be isolated in the depth image by using region based segmentation, or similar approaches. The depth image is converted into a 3D point cloud, using the intrinsic parameters of the depth camera. The point cloud, along with the neighborhood information from the depth image, is used to generate the 3D mesh by applying a very fast image based meshing approach [9]. The color image is then mapped to the vertices of the mesh, using the extrinsic calibration between the color and depth cameras to create a fully textured 3D mesh. The color image is cropped to the relevant size, and compressed to JPEG on the GPU, to reduce the data size for faster transmission.

Even while rendering on 3D displays, depending on the RGB-D camera arrangement and the position of the virtual camera, only a small subset of meshes generated from certain cameras are required to be rendered. A view based camera selection approach [10] is provided by the framework to select the meshes to be rendered, based on the virtual camera view. For faster performance, only the camera machines providing the selected meshes need to capture, process, and transmit the data. For example, view based rendering achieves better quality rendering, as shown in Figure 3, by adding only 25% (8 ms) to the latency, in comparison to rendering only the front and back meshes. For a given camera setup, the rendering quality can be varied by changing the level of details, and the degree of overlap required for rendering a mesh. Rendering overlapping meshes directly leads to many rendering artifacts. A fragment shader similar to [11], that considers the camera capture, virtual camera, surface normal, and virtual lighting to estimate the color to be rendered, is used to eliminate the artifacts and blend the model inside the scene. The model's interaction with virtual lighting produces shadows, improving user's depth perception.

Fig. 3: Side view of the player mesh using view based rendering (left), and front and back mesh rendering (right).

B. *Interaction*

The framework provides full body interaction using physics colliders. In an i3DTI system a new mesh is created every frame, and estimating mesh based colliders, every frame, will add significant latency. So instead of waiting for the mesh, the skeleton of the user is used to create the colliders. The skeleton is either provided by the camera (MS Kinect V1 and V2), or can be estimated from the depth image. The accuracy of the detected skeleton is susceptible to noise and occlusion. By orienting the camera to see the user completely, the detected skeleton accuracy can be increased. Even though the speed and orientations of the joints, and the corresponding collisions may not be accurate all the time, our observations show that this model is still reliable. Any i3DTI environment requires the person to see the screen, thereby forcing them to face the front camera at all the times. This reduces the possibility of severe occlusion, which results in good skeleton detection.

Fast collision detection is achieved by covering the body with a combination of box and capsule colliders. Box colliders map the chest and abdomen region of the person. All the other

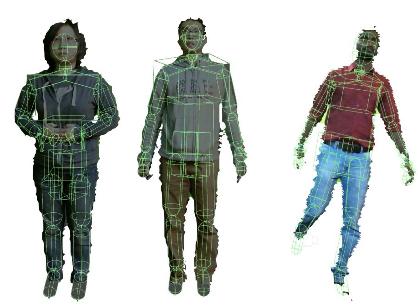


Fig. 4: Skeleton based colliders are shown in green for players of different size.

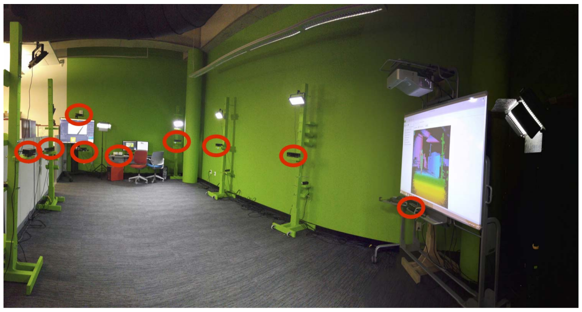


Fig. 5: Play area setup, with multiple MS Kinect V2 cameras (marked in red) for capture, TV, and projector for rendering.

parts, including the head, are represented by a capsule collider. Since the skeleton represents the medial of the person's mesh, it is possible to extract a reasonably good estimate for the size of various parts of the body using just the skeleton. The human body is largely symmetrical, and many parts of the body are proportional to each other. For any given skeleton, a joint association based model is used to estimate the size and orientation of each collider. The entire human model is made collision capable in microseconds, without needing any mesh or point cloud information, as shown in Figure 4.

IV. The Game

The combination of the rendering and interaction pipeline of the i3DTI framework makes it possible to create very engaging games with good quality graphics. To study the preferences of the user, in terms of VQ, level of interaction and better immersion, the game needs to cater specifically to highlight the benefits and minimize the limitations of the i3DTI technology.

A. Game Requirements

Like video games, an i3DTI game needs to be intuitive, easy to play and engaging. The fundamental benefit of an i3DTI game is the ability to embed the user directly into the game. This results in a one-to-one correspondence between the user's expressions, movements, actions, etc., in the real and virtual world. The capture area in the real world is limited in size, as shown in Figure 5. To keep the game realistic, and maintain the direct correspondence, the motion of the player in the game should also be limited. Occlusion, while capturing the user, results in both an inaccurate skeleton and a reconstruction with holes and other artifacts. To reduce the amount of self occlusion, the actions required to play the game should position the user in poses that can be captured clearly by all the cameras. Using external input devices causes visual occlusion, and provides a less engaging experience.

B. Penalty Shootout

Many different games can satisfy the requirements listed in Section IV-A. The penalty shootout situation in soccer meets all of the criteria, and also allows for individual practice play. The penalty game is played by two players, positioned as the striker or the goalkeeper. The striker is supposed to send the ball through the goal post by kicking or using any other part of the body, except their hands. The goalkeeper is expected to stop the ball from going into the net, using their entire body. The striker is positioned at the penalty spot and the goalkeeper is on the goal line, as shown in Figure 1.

The player is provided a default third person point of view at the start of the game, and can change the virtual cameras orientation and location to match their preference and convert to a first person view. To improve the game play, the player is rendered translucently, allowing them to see the ball and other objects through their body. The texture of the player model is altered based on the virtual scene lighting, to provide realistic rendering. An accurate shadow of the player is rendered by positioning multiple virtual light sources. The shadows in the scene enhance the user's capability to track the ball and other objects in the game.

C. Mini Games

While penalty shootout can be played and enjoyed without prior knowledge or skill, possessing good skills enhances the two player gaming experience. The mini games allow the players to practice alone, to improve the necessary skills needed to play the penalty shootout game. The mini games were equally challenging and required slightly different skills. They also allowed for an independent assessment of the player's perception about the system, and the skill level of the player. The following three mini games were created, with increasing level of difficulty:

Goalkeeper: The player learns the basic skills necessary to be a goalkeeper, like moving in the virtual world, ball tracking, and blocking the ball with the body. The ball is placed at the penalty spot and the ball's projected target region is shown to the player a second before the ball is kicked, as shown in Figure 6.

Targeted shooting: This game is designed to increase the precision of basic skills that were developed in the previous goalkeeper mini game. The player is supposed to kick a static ball, positioned at the penalty spot, as shown in Figure 6, into the net to hit the targets.

Moving ball shooting: The ball is placed in the field of view of the player and passed towards them. The player is expected to either stop and kick, or directly deflect the ball, to hit the targets inside the goal. To be successful the player needs to be able to track the ball, move appropriately, time and direct the kick accurately.

The inherent system lag, 3D reconstruction quality, immersion, collision accuracy, real to virtual world correspondence,

Fig. 6: Goalkeeper, with the circle showing the expected ball position (left), and the striker with practice targets (right).

(a) 3D TV third person view

(b) Oculus Rift first person view

Fig. 7: One frame each as shown to the participant in the two player Penalty Shootout game using third person rendering on 3D TV (left) and first person rendering on Oculus Rift (right).

rendering quality, etc. are important factors affecting in-game player performance. Playing the mini games forces the players to be engaged and adapt to the inherent deficiencies of the system. This helps to ensure the player's ability to evaluate the system confidently.

D. Implementation

The i3DTI setup consisted of a total of 12 MS Kinect V2 cameras that were positioned around the room, as shown in Figure 5. Each of the cameras are connected to a camera machine (Intel Xeon 3.0 ghz processor, 18GB RAM and Nvidia Quadro 4000 graphics). For the two player penalty game, the capture area was divided equally into two sites, each containing 6 cameras to capture the user. At each site, 3 cameras were placed in front of and behind the user, so that each side of the user was captured by 2 cameras. The cameras are pointed downward to capture the person from head to toe. Each site contained a rendering machine (Intel i7 2.4ghz processor, 32GB RAM and Nvidia GTX 970 graphics) that was connected to either a 3D TV, projector or Oculus Rift. All of the machines are connected to a local switch, using gigabit networking for the camera machines, and 10GbE networking for the rendering machines.

The i3DTI system is implemented using massively parallel architectures in C++ and CUDA, to minimize system latency. The work load is also distributed across a cluster of machines used for both capture and rendering. The camera machines reduce noise using a median filter, segment the user and generate a 3D mesh from the depth image. The camera machines also generate texture mappings between the color image and the 3D user mesh, crop the relevant regions, and compress the image using JPEG. The penalty shootout and related mini games were developed in C# on the Unity3D gaming engine. TCP Sockets were used to communicate between the i3DTI framework and the game. The MS Kinect positioned in front of the display was used to track the skeleton of the user. The game was rendered to the user in stereoscopic 3D, on either the 3D TV, projector or Oculus Rift depending on the intended third/first person use as well as the site. The system clocks of all the machines in the i3DTI setup are synchronized using NTP, to within a millisecond. The system latency is measured as the time between camera capture and the corresponding mesh rendering.

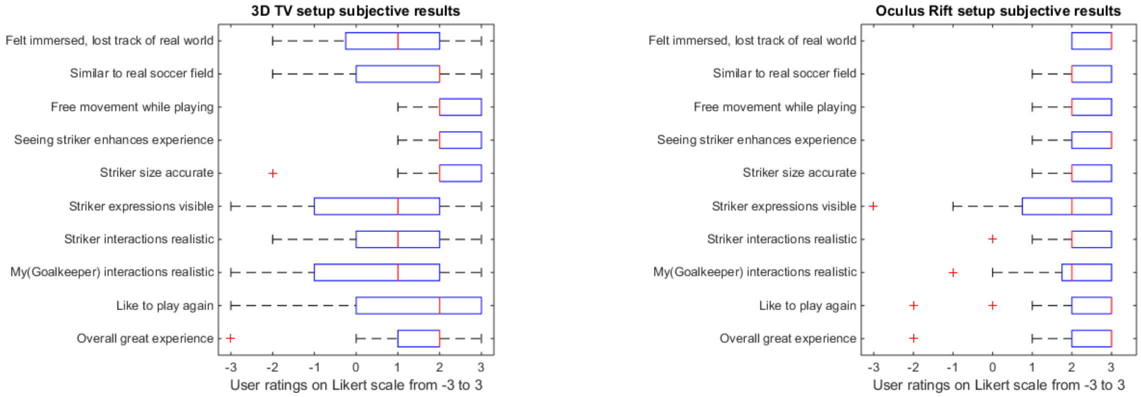
V. User Study

To answer the question of what helps the users have better game experience - higher VQ, better interaction or better immersion, a multiple stage user study was performed using the penalty shootout game. Two different scenarios were set up for performing the user study.

- **Penalty shootout:** A two player penalty shootout study was designed to understand the impact of immersion on the overall QoE. Immersion quality was compared using different rendering hardware devices used - 3D TV for third person setup and head mounted display - Oculus Rift for first person view. The VQ of the 3D models were kept constant and were set so that the interactions in the game were fast and realistic.
- **Mini-games:** The mini-games study was designed to answer the higher VQ or better interaction question in third person view on a 3D TV, keeping the immersion aspect of the game constant. Two different scenarios were set up for this study. The first setup was with Optimized Visual Quality that had high VQ with more cameras. The other was with Optimized Interaction Quality that had less number of cameras sacrificing the VQ but improving the in-game interactions.

A. Penalty Shootout Study

The penalty shootout game answers the question about the impact of better immersion on the user game experience. The VQ and interaction aspects of the game are kept consistent so as to not bias the immersion study. Two different scenarios were set up for this penalty shootout game. Both the scenarios had two remote users present in the game at the same time. It was setup using two adjacent sites connected by a ultra high speed network. Since the setup was in a controlled lap environment we could achieve high network transfer speeds and low latency, allowing for high VQ of the 3D models generated. At the same time, the user's interactions in the game were kept fast and realistic. The game was implemented to render 2 meshes captured by 2 cameras at the same time for each person. At one site, the participant would be present and s/he would be playing the penalty game as the goalkeeper. On the other site, one of the authors was present acting as the striker.



(a) User ratings for 3D TV setup

(b) User ratings for Oculus Rift setup

Fig. 8: Box plots of user ratings for the two setups of the Penalty Shootout game.

The first setup was on a 3D TV, in which the participants would play the game in third-person view. On a 3D TV it was difficult for the users to perceive their position and play the game in first-person view, hence the third-person view was adopted. They would see a transparent back-side view of themselves in the game. They would also see the striker standing at some distance from the goal and in front of the soccer ball. The transparency aspect for the goalkeeper was added to allow the participants to see the motion of the ball even if their 3D model would be in the line of sight of the virtual camera and the ball. One such frame from this setup is as shown in Figure 7a. The second setup was on Oculus Rift, which is a head mounted display device. The participant would in a first person view, with the virtual camera attached that would follow the head movement and show parts of the scene, similar to how the person would see through their own eyes if they were really present there. In this setup the participant would only see the striker and the ball while playing the game, as shown in Figure 7b.

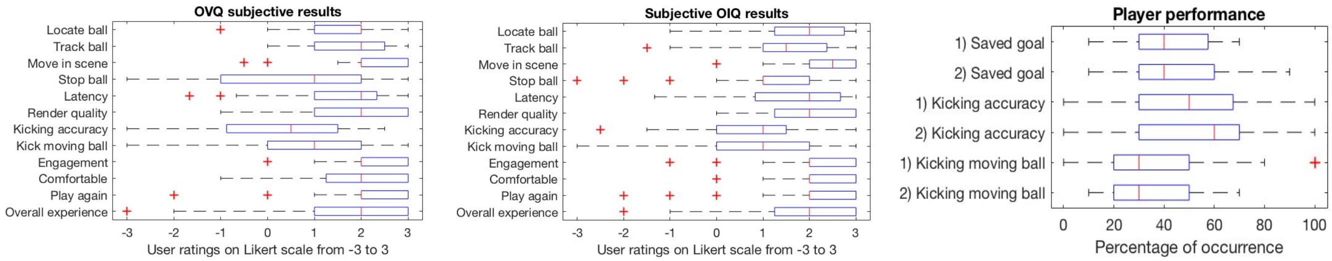
Task of the game, for the participant, was to save/stop the soccer ball, which was kicked by the striker, from going inside the goal area. In each of the two scenarios, the participant had to try and save as many goals as possible from the 10 kicks made by the striker. Before starting the game, the participants were explained about the game and the two setups that were to be performed. They were also shown the 10 different questions which were supposed to filled in by the participant after the study. Since this study was performed with general users which have minimal to no knowledge about 3DTI systems or QoE factors in such systems, they were made aware of the questions beforehand. This would allow them to be aware of the aspects of the game to rate while actually playing the game. The participants rated the 10 different questions that were asked, per scenario, using Likert scale ranging from -3 to 3, with -3 being strongly disagree and 3 being strongly agree. There were few questions, which required the user to evaluate the VQ of the person present in the scene. They were explicitly explained to rate the VQ of the striker and not the goalkeeper. We assigned only one person to be the striker so as to have consistency while answering those questions about VQ. All the participants first played TV setup followed by Oculus Rift setup so as to maintain consistency in the study.

The study was performed by 25 volunteers, 12 male and 13 female participants with their age ranging from 22 to 29. Some basic questions were asked to the participants in terms of their experience with computer games, 3D presence, virtual reality and soccer. Based on their answers, the group consisted of 18 computer gamers, 7 regular soccer players, and 15 VR users. Only 5 of the participants indicated that they had neither experienced or seen 3D presence in action.

Based on the study results, shown in Figure 8a & 8b, almost all users loved the game and would like to play the game again no matter which scenario. They also felt that both the setups had ample space to allow them to move in the real world to play the game properly. All of the users strongly agreed that the presence of the other player, striker, in the game enhanced the user experience. Comparing both the result graphs, the users felt that Oculus Rift provided with a better view of the virtual game and made it look more realistic similar to a real soccer field, rather than on the 3D TV. The striker appearance, expressions as well as interactions were perceived to be better in the Oculus Rift first person setup. From the graph it is clear that almost all of the users felt that the Oculus Rift setup made them feel more immersed in the game, and preferred this setup over the 3D TV setup. At the same time, they felt that their own interactions in the game were better in first person rather than third person. In fact, this was reflected in their goal saving scores. On an average 5 goals were saved, out of 10, in the 3D TV setup. Whereas the average number of goals saved in the Oculus Rift setup were 8, which is a clear improvement in the user's game performance. Overall, the penalty shootout game provided a great experience in both the scenarios, with the first person Oculus Rift setup having better immersion, performance and game play experience than the 3D TV setup.

B. Mini Games Study

The penalty shootout study was setup to evaluate how much impact does immersion have on the user game experience. The study was possible in a controlled lab setup at low latency, high VQ and interaction. We studied the latency logs in a realistic i3DTI system and found that the average latency for local site was 43 ms, and was 67 ms for remote player. There



(a) User ratings for OVQ study (b) User ratings for OIQ study (c) Player performance - OVQ(1) & OIQ(2).

Fig. 9: From left to right: Box plots of user ratings in OVQ and OIQ, and quantitative statistics of player performance.

were some jitters, up to 30 ms, that were noticed during a couple of sessions. Considering at least two sites and having high VQ, the latency of the system increases to about 100 ms leading to slower interactions in the system. In fact, to obtain high VQ more cameras would be required which leads to the latency going of the roof. Compromise has to be made on the VQ to obtain better interaction. The mini games study focuses on the impact of VQ and interaction on the overall QoE in an i3DTI system.

Two versions of the mini games were created 1) A view based rendering version for Optimized Visual Quality (OVQ) and 2) A low latency version for Optimized Interaction Quality (OIQ). The participants first finished all of the OVQ mini games, followed by all of the OIQ mini games. The player's performance and system latency were monitored during each of the games played. A holistic view of each game played is provided by the subjective user evaluations, quantitative player's performance, and qualitative system latency information. All of the user study participants played the games in the order described in Section IV-C. Each participant was allowed to play a minimum of 10 turns, or more if requested by the user. Players were allowed to customize the virtual camera view using verbal instructions while playing the game. After playing each mini game, the user rated various aspects of their game play experience on a 7 point Likert scale, ranging from strongly agree (3) to strongly disagree (-3).

A total of 39 unpaid volunteers (21 male, 18 female), with varying backgrounds in the age group of 20 to 25, were recruited to be part of this user study. All the participants were given an initial survey to determine their experience levels. Based on their replies, the study group consisted of 26 computer gamers, 15 regular soccer players, and 20 VR users. Only 9 of the participants indicated that they had neither experienced or seen 3D presence in action.

1) Optimized Visual Quality: The OVQ version of the games is configured to use view based rendering for displaying the user. The view based rendering of the user is captured by 6 uniformly distributed cameras, results in, at most, 3 camera meshes being rendered for most viewing angles. The user study questions were focused on the user's perception of their own performance and system aspects. Many of the players with no experience in gaming, VR, soccer, or 3DTI showed poor technique while playing the games. These players struggled to kick a goal, or even stop a ball heading towards the goal. The user ratings, grouped by question category, are shown in Figure 9a. Based on the user ratings, it is clear that the users are happy with the performance of the system. The lowest user ratings were given to the questions that related to ball interactions. Considering the user group had only a few soccer players, others seem to require a greater amount of time to become acclimated to the system. After each game, the users were asked to rate the responsiveness of the system. In many of the cases, the user responses for the 3 questions varied by as much as 5 points. These did not correlate with the system latencies measured during the same game play. This clearly shows that a few users, with no VR or gaming experience, are not able of quantifying the responsiveness of the system.

2) Optimized Interaction Quality: The OIQ version of the games uses only the meshes from the front and back cameras to render the user. When the player model is viewed from any angle except straight from the front or back, a large hole in the side is visible, as shown in Figure 3. The same set of questions as Section V-B1 were answered by the user after playing the games in OIQ mode. The inexperienced players continued to struggle to perform even the simplest of tasks, like kicking the stationary ball. The difference between the average system latencies, of OVQ (39 ms) and OIQ (31 ms) versions of the games, was only 8ms. Compared to the ratings of the OVQ version, the OIQ version's user ratings, showed in Figure 9b, displayed a clear improvement in all of the categories, except tracking the ball. The OIQ game versions use only two cameras, leading to visible holes in the side of the body. Still there was an overall increase of about 0.5 points for the user rendering quality. Considering there were no changes in physics or interactions between the OVQ and OIQ game implementations, the improved user ratings for the ball interaction questions can be attributed to the user's ability to adapt to the system. Based purely on the user perception captured in the user study, it is clear that the users prefer faster interactions over VQ while playing an engaging game.

If only the subjective views of the users are considered, then most of them were able to play the game very well. However, the quantitative results of player performance provide a totally different outlook. There was no significant difference in the overall player performance between OVQ and OIQ implementations, as seen in Figure 9c. The perception of improvement in kicking accuracy, shown by the subjective user ratings, turn out to be baseless. There is no evidence of an increase in kicking accuracy between the OVQ and OIQ implementations. This shows that the players engaged in playing the game seem to lose a sense of perception in the real world. In situations where the user played longer, the player performance showed significant improvement, between the OVQ game play and the corresponding OIQ game play. A few users with no prior gaming, VR, or soccer experience seemed to struggle even to kick the ball. Players having difficultly locating the static ball, struggled in both situations and rated the system poorly. The skeleton tracking of the foot is dependent on the footwear, and kicks made by users with shiny shoes were going in unintended directions. Overall, despite the low user ratings given for the kicking accuracy, some users with soccer experience were able to kick the ball accurately, while some with no gaming or soccer experience struggled.

The difference between the subjective user ratings, quantitative player performance, and the system latencies clearly indicates that the user is completely immersed in the game. With the focus primarily on improving their skills and better game play, the player loses a sense of time and develops an elevated sense of performance. This trend can be clearly seen by comparing the results of the OVQ and OIQ user studies, especially in the case of mesh rendering quality, where a mesh with holes on the side is rated higher than a mesh without it.

VI. Conclusion

A multiple camera i3DTI application was made possible by the low latency i3DTI framework implementation. The distributed i3DTI frameworks GPU based implementation provides view based processing, and skeleton based full body interactions, at very low latency. The impact of VQ, faster responsiveness and better immersion on a player's perception and experience in an i3DTI game were studied. This also allowed to further reduce the system latency by reducing the VQ. A penalty shootout game setup at two sites, with a total of 12 Kinect V2 cameras, was created using the i3DTI framework. The users found the penalty shootout game to be visually appealing, intuitive, engaging, and highly entertaining. Using the two player penalty shootout game, the impact of immersion on the QoE was clearly noted. Users preferred to have high immersion in the system which was obtained using first person view on Oculus Rift compared to the third person view on a 3D TV. Goalkeeper, targeted shooting, and moving ball shooting single player mini games were created to train the user. An OVQ and OIQ version of the single player mini games were developed to study the QoE in terms of user behavior in high VQ and high responsiveness situations. Similar to the penalty shootout game, the users were satisfied with the experience in both of the OVQ and OIQ situations. Players rated the VQ, of both of the OVQ and OIQ setups, similarly despite a clearly visible visual disparity between the two renderings. This clearly reiterates that interaction aspect of the game is an important factor for QoE along with VQ. An i3DTI system, that considers both the viewing angle and level of interaction for selecting cameras to be processed, transmitted, and rendered, can vastly improve the user's perception of VQ while maintaining user engagement. The QoE studies presented bring forth an important conclusion that having high VQ in such systems is not the only aspect, other aspects such as better immersion and realistic interactions do matter a lot, if not more than VQ itself.

Acknowledgement

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 1012975 and US Army Research Office (ARO) Grant W911NF-17-1-0299. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF and ARO.

References

[1] W. Wu, A. Arefin, Z. Huang, P. Agarwal, S. Shi, R. Rivas, and K. Nahrstedt, “”i'm the jedi!” - a case study of user experience in 3d tele-immersive gaming,” in *Multimedia (ISM), 2010 IEEE International Symposium on*, Dec 2010, pp. 220–227.

[2] S. Raghuraman, K. Venkatraman, Z. Wang, J. Wu, J. Clements, R. Lotfian, B. Prabhakaran, X. Guo, R. Jafari, and K. Nahrstedt, “Immersive multiplayer tennis with microsoft kinect and body sensor networks,” in *Proceedings of the 20th ACM Multimedia Conference, MM '12, Nara, Japan, October 29 - November 02, 2012*, 2012, pp. 1481–1484. [Online]. Available: http://doi.acm.org/10.1145/2393347.2396526

[3] K. Venkatraman, S. Raghuraman, and B. Prabhakaran, “Demo paper: Demonstration of a 3d tele-immersion framework,” in *2013 IEEE International Conference on Multimedia and Expo Workshops, San Jose, CA, USA, July 15-19, 2013*, 2013, pp. 1–2. [Online]. Available: https://doi.org/10.1109/ICMEW.2013.6618232

[4] K. Desai, K. Bahirat, S. Ramalingam, B. Prabhakaran, T. Annaswamy, and U. E. Makris, “Augmented reality-based exergames for rehabilitation,” in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16, 2016, pp. 22:1–22:10.

[5] K. Desai, U. H. H. Belmonte, R. Jin, B. Prabhakaran, P. Diehl, V. A. Ramirez, V. Johnson, and M. Gans, “Experiences with multi-modal collaborative virtual laboratory (mmcvl),” in *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, April 2017, pp. 376–383.

[6] W. Wu, A. Arefin, G. Kurillo, P. Agarwal, K. Nahrstedt, and R. Bajcsy, “Color-plus-depth level-of-detail in 3d tele-immersive video: a psychophysical approach,” in *Proceedings of the 19th ACM international conference on Multimedia*, ser. MM '11. New York, NY, USA: ACM, 2011, pp. 13–22. [Online]. Available: http://doi.acm.org/10.1145/2072298.2072302

[7] H. C. Miles, S. R. Pop, S. J. Watt, G. P. Lawrence, and N. W. John, “A review of virtual environments for training in ball sports,” *Computers & Graphics*, vol. 36, no. 6, pp. 714 – 726, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849312000957

[8] G. Kurillo and R. Bajcsy, “3d teleimmersion for collaboration and interaction of geographically distributed users,” *Virtual Reality*, vol. 17, no. 1, pp. 29–43, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10055-012-0217-2

[9] S. Raghuraman and B. Prabhakaran, “Distortion score based pose selection for 3d tele-immersion,” in *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '15. New York, NY, USA: ACM, 2015, pp. 227–236. [Online]. Available: http://doi.acm.org/10.1145/2821592.2821600

[10] K. Pulli, H. Hoppe, M. Cohen, L. Shapiro, T. Duchamp, and W. Stuetzle, “View-based rendering: Visualizing real objects from scanned range and color data,” in *Rendering Techniques 97*, ser. Eurographics, J. Dorsey and P. Slusallek, Eds. Springer Vienna, 1997, pp. 23–34. [Online]. Available: http://dx.doi.org/10.1007/978-3-7091-6858-5_3

[11] R. Vasudevan, G. Kurillo, E. Lobaton, T. Bernardin, O. Kreylos, R. Bajcsy, and K. Nahrstedt, “High-quality visualization for geographically distributed 3-d teleimmersive applications,” *Multimedia, IEEE Transactions on*, vol. 13, no. 3, pp. 573–584, June 2011.