FreeViewer: An Intelligent Director for 3D Tele-Immersion System

Zhenhuan Gao Shannon Chen Klara Nahrstedt
Department of Computer Science
University of Illinois at Urbana-Champaign
{zgao11, cchen116, klara}@illinois.edu

ABSTRACT

This paper proposes FreeViewer, a 3D Tele-Immersion view-control system that allows viewers to see arbitrary side of the performer by intelligently choosing the streams of a subset of cameras and changing the point of view in a 3D virtual space. The view changing is actuated by the change of the sensor data from wearable devices (eg. Google Glass, smartphone) on the performer able to monitor the current orientation.

Categories and Subject Descriptors

J.7 [**COMPUTERS IN OTHER SYSTEMS**]: Real time

Keywords

3D tele-immersion; multi-modality; free view

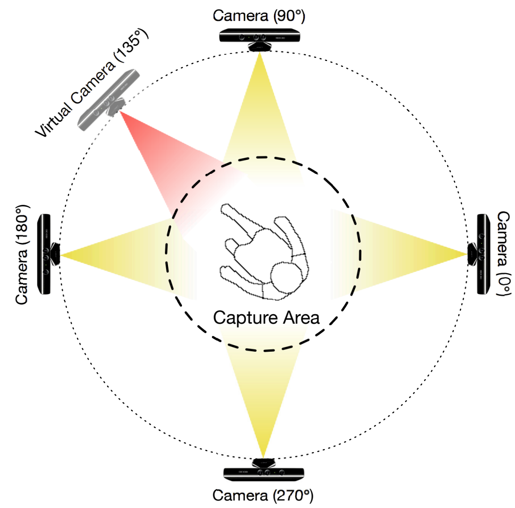

Figure 1: The bird view of FreeViewer deployment

1. INTRODUCTION

3D Tele-immersion (3DTI) system enables remote users to view the local scene and even to interact with local users in a virtual 3D space, which has promoted lots of potential applications such as remote therapy [1]. In general, a 3DTI system consists of several sites distributed at different physical locations. All the sites transmit the 3D data captured by multiple 3D cameras over either a dedicated network or a shared network such as Internet. After each site gathers the data from all the other sites, it renders everything in a common virtual space in a real-time manner.

In a traditional 2D live video streaming system, the interactivity of end users, choosing a specified viewpoint, has been crippled by the fact that they can only choose to see the physical scene captured by a physical camera, but not between two physical cameras. However, through depth estimation and interpolation among multiple 3D video streams captured from different angles, the 3DTI system makes it possible rendering a 3D space where the viewers can view physical scene from arbitrary viewpoint as if there existed a virtual camera capturing at that viewpoint with the same view angle.

One possible solution to realize free viewpoint (FV) viewing of a physical scene is to create a complete 3D model by aggregating the 3D streams from the cameras all of which can cover 360 degree view. The viewing user is then able to choose arbitrary viewpoint to see any side of this complete 3D model. However, creation and delivery of such a complete 3D object makes bandwidth and computation requirements inevitably high. Hence, we need to look for alternative solutions for FV viewing to decrease the computational and networking requirements. For example, in the 3DTI system, the back side of the model, which is invisible to the viewer, is redundant. The computation related to the data of the back could be avoided to reduce latency.

In this paper, to tackle the FV viewing problem, we propose FreeViewer, a 3DTI view-control system at the physical scene that supports free viewpoint with low requirements on computation power and bandwidth and is suitable to assist the real-time peer-to-peer 3D content distribution among large scale viewers. Figure 1 illustrates an example of the deployment of FreeViewer. There are 4 3D cameras deployed in the area with the same angular difference and the same distance from the center of the capturing area. A performer stays in the capturing area with a varying orientation. The objective of FreeViewer is no matter which direction the

MM'14, November 3–7, 2014, Orlando, Florida, USA.
ACM 978-1-4503-3063-3/14/11.
http://dx.doi.org/10.1145/2647868.2654873 .

performer is facing, it always renders the scene of the performer's front side as if there was a virtual camera facing the performer.

2. SYSTEM OVERVIEW

This section gives an overview of the FreeViewer system architecture (Figure 2) including several important functional services.

2.1 3D View Tracking Service

3D View Tracking is an advanced service that performs cooperation between wearable device and orientation monitoring module to capture user's head direction. Traditional user interaction solutions for 3DTI systems are often naive and simply relying on standard iuput devices(mouse, keyboard) to control visual devices and receive visual feedback. We present a user interaction solution based on wearable devices. That is, we use Google Glass as the input interaction device to enable tracking the rotational movement of the user's head. The tracking service allows users to interact with the system in a natural way without any additional manual operation such as "press some keys" or "drag the mouse". We choose Google Glass because of its portability and easiness to estimate the facing orientation.

The orientation monitoring module also maintains a connection with the Google Glass to receive the sensor data and then feeds them to capturing service and rendering service.

2.2 Capturing Service

As previously stated, in a 3D virtual space at a specified viewpoint, a viewer can only see one side of an object which can be covered by at most two 3D cameras. This means in order to render the scene from any viewpoint at any time, our system only needs the streams from two cameras. Given the orientation returned by the Google Glass, the capturing service decides the streams from which cameras to acquire according to the known physical deployment of the cameras. Then this service synchronizes the color and depth streams from both of the cameras chosen and sends the frames to the processing service to create a partial 3D model.

In fact, it not only reduces the bandwidth consumption of the network resources but also practically relieves the interference effect [2] between different 3D cameras (e.g., Kinect).

2.3 3D Point Cloud Processing Service

The 3D cloud point processing service aims to use the color stream and the depth streams from the 3D cameras to generate two 3D model (Point Clouds [3]) and to merge them in to a complete model so that the intersecting areas between them could overlap perfectly. The problem of such consistent alignment is known as registration.

However, the most famous registration algorithm, the iterative closest point algorithm, is too slow to support real-time processing. Fortunately, in our deployment, we already know the physical positions and orientations of all cameras around the captured physical scene because the cameras are fixed prior to filming and the capturing of actions in front of the cameras. Thus, we adopt the pre-configuration and calibration technique to obtain the coordinates-transformation relationship between a main camera and the other cameras. Then after the system starts to run, the processing service transforms all 3D point clouds from the original cameras' coordinate system to the coordinate system of the main camera, which aligns them properly.

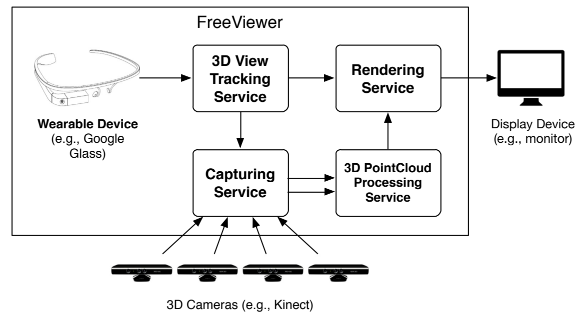

Figure 2: FreeViewer Architecture

Finally, this service delivers the merged point cloud, which is also a partial 3D model, to the rendering service.

2.4 Rendering Service

The rendering service takes two inputs, the performer's current orientation, coming from the 3D view-tracking service, and the correspondent partial 3D model in terms of merged point cloud. Based on the user's orientation and the geometric relationship between the deployed cameras, the rendering service calculates the virtual camera's viewpoint and view angle. Then it uses standard OpenGL API to render the synthesized scene on the display device. Also, in order to provide truly immersive visual feedback, the display device can also be a head-mounted display or even another Google Glass with OpenGL ES support.

3. ACKNOWLEDGEMENTS

This material is based upon work supported by NSF Grant CNS10-12194.

4. CONCLUSIONS

The presented demo shows the attractiveness of the FreeViewer system in terms of view-changing capabilities for a user via her wearable device such as the Google Glass in multi-camera immersive environment. The view-tracking technique is based on the sensor data from wearable devices. Compared to complete 3D model reconstruction, FreeViewer can achieve the same FV viewing and the same 3D video quality with less computation and network bandwidth, which is highly favorable under resource-limited scenarios.

5. REFERENCES

[1] Klara Nahrstedt. 3d teleimmersion for remote injury assessment. In *Proceedings of the 2012 International Workshop on Socially-aware Multimedia*, SAM '12, pages 21–24, New York, NY, USA, 2012. ACM.

[2] D. Alex Butler, Shahram Izadi, Otmar Hilliges, David Molyneaux, Steve Hodges, and David Kim. Shake'n'sense: Reducing interference for overlapping structured light depth cameras. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1933–1936, New York, NY, USA, 2012. ACM.

[3] Point cloud library. `http://pointclouds.org/`.