GPU-Based View Synthesis Using an Orbital Reconstruction Frustum

Greg Welch
University *of* North Carolina
at Chapel Hill
Chapel Hill, NC 27599 USA
welch@cs.unc.edu

Hua Yang
University *of* North Carolina
at Chapel Hill
Chapel Hill, NC 27599 USA
yanghua@cs.unc.edu

Andrei State
University *of* North Carolina
at Chapel Hill
Chapel Hill, NC 27599 USA
andrei@cs.unc.edu

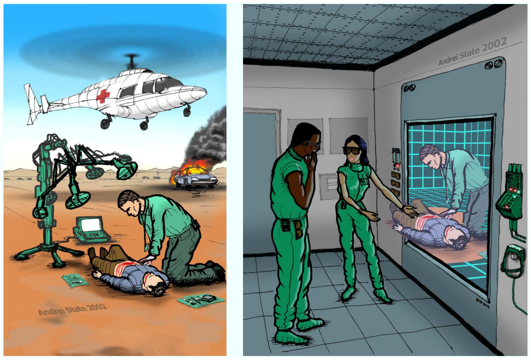

Figure 1: Conceptual sketches of some three-dimensional (3D) medical telepresence paradigms. Left: a portable multi-camera unit. Right: hospital-based 3D viewing. (Sketches by Andrei State.)

1. INTRODUCTION

Two-dimensional (2D) video-based telemedical consultation has been explored widely in the past 15–20 years. Two issues that seem to arise in most relevant case studies are the difficulty associated with obtaining the desired 2D camera views, and poor depth perception. To address these problems we are exploring the use of a small array of cameras to reconstruct a real-time, on-line 3D computer model of the real environment and events. We call this 3D medical consultation (3DMC). The idea is to give remote users what amounts to an infinite set of stereoscopic viewpoints, simultaneously addressing the visibility and depth perception problems associated with 2D video. The idea is illustrated in Figure 1.

Figure 2: 3DMC prototype. Left: a camera-lighting array with eight Firewire cameras and high-frequency area lights. Middle: transportable rack with five high-performance GPU-equipped servers. Right: remote 3D viewing station.

Our current 3DMC prototype consists of a camera-lighting array, a transportable compute/rendering cluster, and a head-tracked display. See Figure 2. The camera-lighting array consists of eight 640 × 480 resolution digital (IEEE 1394a) color cameras from Point Grey Research. The cameras are positioned so their visual fields converge on the same region of interest on the patient. See also Figure 3.

The compute/rendering cluster (middle in Figure 2) consists of five dual-processor PCs mounted in a transportable shock-resistent rack case. Four PCs function as camera servers, converting the raw Bayer pattern images into RGB images and compressing them using JPEG then forwarding the compressed video streams via dedicated gigabit Ethernet to the 5^{th} PC. The 5^{th} PC then decompresses the video streams, loading the color images into texture memory of a Radeon X800XT 256MB AGP8X graphics card for GPU-based view synthesis.

2. GPU-BASED VIEW SYNTHESIS

To achieve a visual sense of 3D telepresence we use a novel approach called *View-dependent Pixel Coloring* (VDPC) [1]. VDPC is a hybrid image-based and geometric approach that estimates the *most likely color* for every pixel of an image that would be seen from some *desired viewpoint*, while simultaneously estimating a 3D model of the scene. By taking into account object occlusions, surface geometry and materials, and lighting effects, VDPC can produce results where other methods fail—in the presence of textureless regions and specular highlights—conditions that are common in medical scenes.

We use the graphics hardware on the 5^{th} PC to perform the 3D reconstruction very quickly as the images arrive from the four camera server PCs. The basic idea is to use the graphics hardware to rapidly render the camera images onto a series of virtual (computer graphics) planes swept through the scene, searching in parallel for the best color matches (least variance) at a dense set of points on the planes [2]. The result is a relatively dense depth map that we can then render, again using the graphics hardware.

3. ORBITAL RECON. FRUSTUM

Our 3DMC prototype reconstructs the scene surface using plane sweeping throughout a synthetic *reconstruction frustum*. The pyramid shaped reconstruction frustum is discretized into parallel planes perpendicular to the major axis of the reconstruction frustum. The planes are partitioned

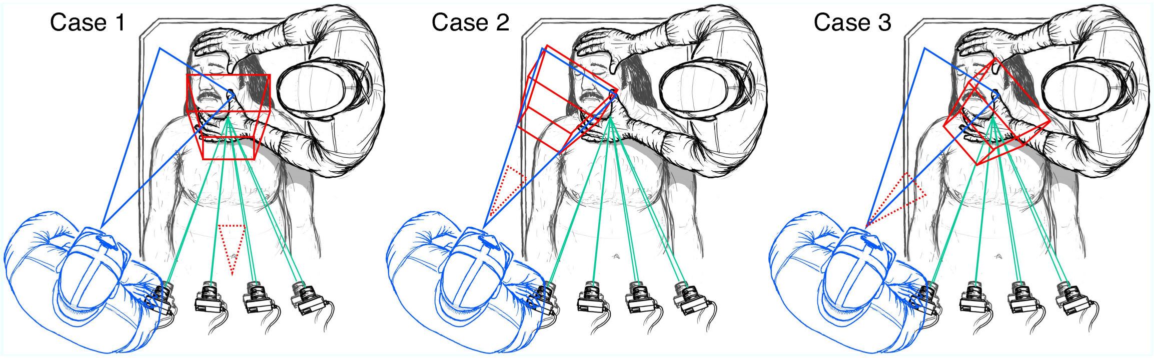

Figure 3: The three cases corresponding to Table 1. The remote viewer (consultant) is on the left. The solid green lines from the cameras indicate their convergence point. The solid and dashed red lines indicate the reconstruction frustum and origin, respectively. The solid blue lines indicate the the gaze of the remote viewer. (Sketches by Andrei State.)

into cells corresponding to image pixels for the desired view. Using the graphics hardware we project the camera images onto these planes. If there is a visible (unoccluded) surface at a cell of a plane, the projected images should have consistent color at that cell. Thus for each pixel on the synthetic image plane we compute the mean and variance of the projected colors of all the corresponding cells while sweeping, and choose the distance and mean color with minimum variance. This reconstructs the scene surface as a set of planar patches that are inside the intersecting volume of the reconstruction and camera frusta, and perpendicular to the major axis of the reconstruction frustum.

Classic plane sweeping methods use a fixed reconstruction volume defined in world coordinates. See Case 1 in Figure 3 and Table 1. For light field configurations where multiple cameras are pointing in similar directions, the reconstruction viewpoint and direction are typically set to the average of the cameras. This achieves maximum overlap between the reconstruction and *camera* frusta, but decouples the reconstruction and *viewing* frusta. This may cause disturbing artifacts in the synthesized images, such as gaps between adjacent patches at different depths. In addition, scene point visibility changes as the viewpoint changes.

On the other hand, one may consider “attaching” the reconstruction volume to the view coordinate frame. See Case 2 in Figure 3 and Table 1. In this case, the reconstruction frustum is aligned with the *viewing* frustum but decoupled from the *camera* frusta. This can result in disturbing dynamic artifacts when the viewer gazes away from the reconstruction frustum.

Table 1: Reconstruction frustum position (POS) and orientation (ORI) with respect to world and view coordinate frames.

Case	WRT *World* Frame	WRT *View* Frame
1	Static POS & ORI	Dynamic POS & ORI
2	Dynamic POS & ORI	Static POS & ORI
3	Static POS Dynamic ORI	Static ORI Dynamic POS

Our 3DMC prototype uses a hybrid *orbital reconstruction frustum* configuration. See Case 3 in Figure 3 and Table 1. Specifically, we fix the *position* of the reconstruction frustum in the world, but determine its *orientation* dynamically based on the remote viewpoint. In effect, the viewer orbits around a fixed reconstruction that always faces them. This keeps the (bounded) reconstruction volume within the camera frusta and centered about the wound of the patient, while still allowing reconstruction from the perspective of the remote viewer, avoiding the gaps between patches and the problem of visibility inconsistency.

4. FUTURE GPU ARCHITECTURES

GPU hardware is increasingly being used for real-time computer vision. A direct firewire connection (and driver support) between cameras and GPU cards, facilitating direct capture to graphics memory, would significantly reduce overall latency and system complexity. Similarly a direct ethernet connection would facilitate low-latency dynamic parameter changes, such as the current viewing position from a head tracking system. This would benefit many interactive applications.

5. ADDITIONAL AUTHORS

Vincent Noel (UNC-Chapel Hill, `vince@cs.unc.edu`), Adrian Ilie (UNC-Chapel Hill, `adyilie@cs.unc.edu`), Ruigang Yang (University of Kentucky, `ryang@cs.uky.edu`), Marc Pollefeys (UNC-Chapel Hill, `marc@cs.unc.edu`), and Henry Fuchs (UNC-Chapel Hill, `fuchs@cs.unc.edu`).

6. REFERENCES

[1] R. Yang. *View-Dependent Pixel Coloring—A Physically-Based Approach for 2D View Synthesis.* Ph.d. thesis, University of North Carolina at Chapel Hill, 2003.

[2] R. Yang, M. Pollefeys, H. Yang, and G. Welch. A unified approach to real-time, multi-resolution, multi-baseline 2d view synthesis and 3d depth estimation using commodity graphics hardware. *International Journal of Image and Graphics (IJIG)*, 4(4):1–25, 2004.