TransCAIP: Live Transmission of Light Field from a Camera Array to an Integral Photography Display

Yuichi Taguchi*  Takafumi Koike††  Keita Takahashi*  Takeshi Naemura*

*The University of Tokyo  ††Hitachi, Ltd.

(a) (b) (c)

Figure 1: Our real-time light field transmission system consists of (a) an array of 64 cameras that captures multi-view videos of a dynamic scene and (b) an integral photography display with 60 viewing directions that reproduces (c) a full-color and full-parallax autostereoscopic 3D video of the scene.

Executive Summary

TransCAIP provides a real-time 3D visual experience by using an array of 64 cameras and an integral photography display with 60 viewing directions. The live 3D scene in front of the camera array is reproduced by the full-color, full-parallax autostereoscopic display with interactive control of viewing parameters.

1 Vision

Three-dimensional TV is a promising technology for providing a more natural and intuitive perception of 3D scenes than existing two-dimensional TV. In particular, live 3D TV systems, which transmit 3D visual information in real time, could have a significant impact on many applications in communication, broadcasting, and entertainment.

The overall goal of our work is to develop a practical live 3D TV system, TransCAIP, which reproduces a full-color 3D video of a scene with both horizontal and vertical parallax in real time. Our system gives users a perception of observing the 3D scene through a window without requiring them to wear special glasses. We use an array of 64 cameras as the input device and an integral photography display with 60 viewing directions as the output device. The main technical goal is to develop a fast and flexible data conversion method between these asymmetric input and output devices, which runs in real time (more than 5 frames per second) on a single PC with GPGPU techniques.

Real-time light field conversion. To connect the asymmetric input and output devices, our system performs real-time light field conversion between 64 input views of 320×240 pixels captured with the camera array and an integral photography image consisting of 60 views of 256×192 pixels. As shown in Fig. 2, using (a) the 64 input views, we first render (b) 60 novel views corresponding to the viewing directions of the display by using an image-based rendering method [Taguchi et al. 2008], and then arrange the rendered pixels to produce (c) an integral photography image. For generating high-quality novel views, our method estimates a view-dependent per-pixel depth map at each rendering camera viewpoint based on a layered representation. For real-time processing on a single PC, the conversion algorithm is fully implemented on a GPU with GPGPU techniques.

Interactive control of 3D viewing parameters. Our system enhances the users’ 3D visual experience by letting them control viewing parameters of the displayed 3D images interactively. As shown in Fig. 3, in the light field conversion method, the rendering cameras are placed at a regular interval such that their viewing directions converge at the same point. The plane whose depth is equal to that of this point is called the convergence plane. The convergence plane corresponds to the display plane of the integral photography display. Since objects near the display plane are reproduced with a higher resolution than those farther from the plane [Hoshino et al. 1998; Zwicker et al. 2007], our system enables users
Figure 2: Overview of our conversion method. Using (a) 64 input views, the method first renders (b) 60 novel views corresponding to the viewing directions of the integral photography display. The rendered pixels are then arranged to produce (c) an integral photography image. All of the conversion processes are performed on a GPU.

Figure 3: Configuration for rendering 60 views.

to set the plane at a desired position in the target scene (Fig. 4). The position of an object relative to the display plane is also determined by the convergence plane. Moreover, users can control the amount of depth reproduced on the display by changing the interval of the rendering cameras (Fig. 5). Users can also control the location of the part of the scene reproduced on the display by changing the positions and view angles of the rendering cameras (Fig. 6). Note that users can interactively perform the viewing parameter control as a software process without reconfiguring the hardware system.

3 Context

Video-based rendering. Free-viewpoint videos can be generated from multi-view videos captured with a camera array or a lens array by using video-based rendering techniques. Some systems perform offline rendering by using pre-acquired multi-view videos as well as a pre-computed geometry model [Kanade et al. 1997; Zitnick et al. 2004; Wilburn et al. 2005], while the others enable online (real-time) rendering from live multi-view videos [Naemura and Harashima 1999; Schirmacher et al. 2001; Naemura et al. 2002; Yang et al. 2002b; Yang et al. 2002a; Yamamoto et al. 2004; Zhang and Chen 2004]. We have also developed an online video-based rendering system that uses an array of 64 network cameras, estimates a view-dependent per-pixel depth map, and produces a free-viewpoint video in real time [Taguchi et al. 2008]. TransCAIP uses the camera array of that system as the input device, but performs much more complex processes (rendering 60 views and generating an integral photography image) in real time on a single PC.

Autostereoscopic 3D displays. Autostereoscopic 3D displays present 3D images without requiring the viewer to wear special glasses by reproducing different viewpoint images for different viewing directions. Some displays are designed to present different views for 360-degrees directions by rotating an LED array [Yendo et al. 2005] or a screen [Otsuka et al. 2006; Jones et al. 2007]. They are suitable for observing an object within that volume. Meanwhile, others using a lenticular lens or a microlens array are suitable for reproducing an entire scene. Lenticular displays are commercially available but basically provide only 1D parallax (either horizontal or vertical). Integral photography displays, by contrast, provide both horizontal and vertical parallax by using a microlens array. Because such a display needs to reproduce huge number of light rays (the resolution of a view times the number of views), recently developed integral photography displays use multiple projectors [Liao et al. 2005; Yang et al. 2008] or a high-density LCD panel [Okano et al. 1999; Arai et al. 2006; Koike et al. 2007]. TransCAIP uses as its output device the integral photography display reported by Koike et al. [2007], but our light field conversion method is applicable to any kind of integral photography and multi-view 3D displays.

Live 3D TV systems. Okano et al. presented integral-photography-based systems that acquire and display light field in real time [Okano et al. 1999; Arai et al. 2006]. Their systems capture a 3D scene by using a lens array and a high-resolution video camera and
Figure 4: Displayed images obtained using different positions of the convergence plane (top images captured from a left viewpoint and bottom images captured from a right viewpoint). In both cases, the person at the convergence plane is reproduced with the highest spatial resolution of the display and appears at the same position in the display regardless of the user’s viewpoint.

Figure 5: Displayed images obtained using different rendering camera intervals (top images captured from a left viewpoint and bottom images captured from a right viewpoint). (a) When all the rendering cameras are set at the same viewpoint, the display acts as a 2D display (i.e., no parallax). (b) A small interval yields a small parallax (a small amount of depth perception). (c) A large interval yields a large parallax. Note, however, that objects farther from the display plane are reproduced with a lower spatial resolution.

Figure 6: Displayed images obtained using different positions and view angles of the rendering cameras. Note that the display presents autostereoscopic 3D images for any target part of the scene.
present the captured video on an integral photography display. They set a large-aperture convex lens, called depth control lens, in front of the lens array in their capturing systems, so that the real images of objects are formed near the lens array and their 3D images are reproduced around the display plane. The position of the depth control lens determines the depth at which objects appear with the highest resolution. Our system controls this effect by changing the position of the convergence plane as a software process, which provides more flexible control than the hardware reconfiguration that their systems need. Moreover, their systems need symmetric input and output devices (i.e., the number of lenses of the array for capturing is same as that of the array for displaying), whereas our system can use asymmetric input and output devices that have different viewpoint layouts thanks to the conversion method.

Matusik and Pfister [2004] also developed a live 3D TV system using symmetric input and output devices: 16 cameras for the input, and 16 projectors with lenticular screens for the output. Although lenticular screens need multi-view images captured at regularly spaced viewpoints, exactly aligning the input cameras in such a manner is impractical. They therefore correct the misalignment of the camera viewpoints by using image-based rendering with approximating the scene geometry as a single plane. Our system uses image-based rendering not only for the viewpoint correction, but also for converting between the completely different viewpoint layouts of the input and output devices, and for interactively controlling the 3D viewing parameters. Moreover, our system is designed to produce 3D images that have both horizontal and vertical parallax, while their system shows images that have only horizontal parallax. Finally, our system uses a rendering method that estimates a per-pixel depth map, which synthesizes higher-quality images than their single plane rendering.

4 User Experience

In our demonstration, users can observe their own 3D images reproduced on the integral photography display in real time. Moreover, users can interactively control the viewing parameters of the displayed 3D images, as shown in Figs. 4–6. The companion video shows more results of changing the viewing parameters in dynamic scenes. However, all of the image and video results are captured by a single-viewpoint camera and do not fully convey the advantage of our system. As in many other 3D display works, we believe demonstration is the best way to convince users that TransCAIP provides the highest resolution. Our system controls this effect by changing the position of the convergence plane as a software process, which provides more flexible control than the hardware reconfiguration that their systems need. Moreover, their systems need symmetric input and output devices (i.e., the number of lenses of the array for capturing is same as that of the array for displaying), whereas our system can use asymmetric input and output devices that have different viewpoint layouts thanks to the conversion method.

Acknowledgments: We would like to thank Prof. Hiroshi Harashima for valuable discussions.

References


