Improving 3D perception by adding eye-tracking based motion parallax - an ICASSP demo

Dinei A. Florencio and Cha Zhang

Microsoft Research
One Microsoft Way, Redmond, WA 98052
{dinei, chazhang}@microsoft.com

Abstract

Motion parallax is one of the most important 3D clues for human vision perception. The majority of the monitors today display contents regardless of the user's viewpoint. In this demo, we track head/eye position of the viewer based on a web camera, and use that information to render view-dependent contents, adding motion parallax to essentially any display device. In the demo we will allow ICASSP participants to try out the system, and perceive the sensation of depth provided by the technology, which is much cheaper and complementary to stereoscopic displays.

1. Demo Description

Adding 3D perception is desired in a variety of applications [1-3]. Many 3D clues are already routinely incorporated into 2D image or graphics, like shading, occlusion, relative size, etc. In general, the two most important clues missing are stereopsis (i.e., the image presented to the right and left eyes are different), and motion parallax (the image displayed changes with the user position).

Stereopsis can be provided through various stereoscopic displays, which is getting very popular in recent years. However, research studies have shown that motion parallax is equal if not more important than stereo in contributing to 3D perception. Nevertheless, the usage of motion parallax has been mostly limited to lab studies, primary due to the fact that it is generally expensive, and does not produce as an intriguing effect as stereopsis. For instance, most existing systems require the user to wear some type of tracking device, e.g., positioning sensor, LED, goggles, etc. This makes the head tracking easier, but it also makes it more expensive and more inconvenient.

In this demo, we present a new system, where we use head- and eye-tracking technology to add the parallax effect, by rendering the image for the corrected eye position. Note the tracking is performed with a simple web camera, and the user is not required to wear any device for the tracking.

While the technology can also be used for stereoscopic displays, we will demo it on a standard LCD display. In this case we will illustrate the added 3D sensation by adding motion parallax without stereo vision. We call the effect "Monocular 3D". Although not as remarkable as true, stereoscopic displays, the hardware cost is essentially zero (we use just a legacy LCD display and a web cam). This makes the technology very appealing, as it can be applied to essentially all existing displays.

The system includes a standard display, and a camera pointing at the user. The camera is used to find the user's (i.e., viewer's) eyes position. Both the display and camera can be standard off-the shelf components. If a stereo display is available, it can be also used. Similarly, if a depth camera or multiple cameras are available, it may increase the robustness of the eye position estimation.

We also would like to point that this technology is particularly suitable for a demo session. Since the display can only display one image at a time, the algorithm will track only the eyes of one viewer, and the effect can be experienced by only one user at a time.

Finally, we point that the tracking may have many applications, besides the one showcased in this demo and discussed in [4, 5]. Other applications include 3D special effects [6] based on background subtraction [8, 9]. Related work from our group also includes applications to 3D audio [10, 11], room modeling [12, 13], sound source localization [14, 15, 16], and crowdsourcing of subjective quality evaluation [17, 18].

REFERENCES

[1] Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multi-view imaging and 3dtv," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10–21, 2007.

[2] Z. Yang, W. Wu, K. Nahrstedt, G. Kurillo, and R. Bajcsy, "Viewcast: View dissemination and management for multi-party 3d tele-immersive environments," in *ACM Multimedia*, 2007.

[3] H. Baker, D. Tanguay, I. Sobel, D. Gelb, M. Goss, W. Culbertson, and T. Malzbender, "The coliseum immersive teleconferencing system," Tech. Rep., HP Labs, 2002.

[4] D. Florencio and C. Zhang, "Multiview Video Compression and Streaming based on Predicted Viewer Position", ICASSP, 2009.

[5] C. Zhang, D. Florencio, and Z. Zhang, "Improving immersive experiences in telecommunication with motion parallax [applications corner]," IEEE Signal Processing Magazine, 2011.

[6] C. Zhang, Z. Yin, D. Florêncio, "Improving Depth Perception with Motion Parallax and Its Application in Teleconferencing," MMSP, 2009.

[7] C. Zhang and D. Florencio, "Joint Tracking and Multiview Video Compression," VCIP, 2010.

[8] S. Varadarajan, L. Karam, and D. Florencio, "Background Subtraction using Spatio-Temporal Continuities," EUVIP, 2010.

[9] S. Varadarajan, L. Karan, and D. Florencio, "Background Recovery from Video Sequences using Motion Parameters," ICASSP, 2009.

[10] M. Song, C. Zhang, D. Florencio, and H. Kang, "Personal 3D audio system with loudspeakers," ICME, 2010.

[11] M. Song, C. Zhang, D. Florencio, and H. Kang, "Enhancing loudspeaker-based 3D audio with room modeling," MMSP, 2010.

[12] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio, "L1 regularized room modeling with compact microphone arrays," ICASSP, 2010.

[13] F. Ribeiro, D. Florencio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," submitted.

[14] F. Ribeiro, C. Zhang, D. Florencio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," submitted.

[15] Y. Rui, D. Florencio, W. Lam and J. Su, "Sound source localization for circular arrays of directional microphones," ICASSP, 2005.

[16] Y. Rui, D. Florencio, "New direct approaches to sound source localization," ICME, 2003.

[17] F. Ribeiro, D. Florencio, C. Zhang and M. Seltzer, "CrowdMOS: An approach for crowdsourcing mean opinion scores studies," ICASSP, 2011

[18] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowsourcing subjective image quality evaluation," ICIP, 2011.