

Universal capture through stereographic multi-perspective recording and scene reconstruction

Volker Kuchelmeister

UNSW University of New South Wales Sydney, iCinema Centre for Interactive Cinema
Research. kuchel@unsw.edu.au

Abstract. This paper describes a prototype for an installation which combines stereographic video and real-time 3D computer graphic to demonstrate a novel method of documenting dance and other performing art practices through multi-perspective recording, volumetric geometry reconstruction and universal playback. The fidelity and high level of detail in the video imagery is augmented and completed with the voxel representation. Multi-perspective recording in combination with voxelization offers a universal view on a scene. A viewer is not limited to one point-of-view or moment in time, but he can explore and analyze the scene freely and without space or time restrictions. Potential areas of application are in performing arts, in professional sport and the Movie FX industry.

Keywords: Augmented / Mixed Reality, Visual Effects, Virtual Reality, Graphics Techniques, Mixed Media, Interactive Computer Graphics

1 Background

1.1 Double District

The background for this project [7] is a stereo/3D video recording of a dance performance specially choreographed by the renowned Japanese choreographer and dancer Saburo Teshigawara for the installation Double District [6] (Fig.1,2). The six-channel stereo video dance installation is configured in the Re-Actor¹, a 5 m diameter hexagonal rear projected stereo/3D visualization environment. The performance was shot simultaneously with six stereo pairs of high-resolution digital video cameras² from six different points of view (Fig. 3,4). Each of these 3D recordings could then be played in Re-Actor, back-projected using twelve projectors and passive polarized stereo onto its six 2.4x2m back-projection surfaces.

The audience watching this work moves freely around the hexagonal room to view individual screens, or step back to observe up to three screens simultaneously. All six screens show the dancers' movements at the same moment in time but seen from six different points of view, analogous to the architecture of the space within which it is projected. The scaling and virtual 3D location of the dancers is such that they appear as life-size bodies, exactly situated and moving about within the confines of its hexagonal enclosure.

¹ Re-Actor created by Sarah Kenderdine and Jeffrey Shaw. Originally developed for their virtual 3D theater work UNMAKEABLELOVE [12,15], Kenderdine and Shaw also conceived its use as a 3D visualization architecture for the multi-view presentation of live performances.

² 12 x Imperx IPX-2M30G, 1600x1200 pixel resolution, 11.84x8.88mm active image area, digital 8/10/12 bits video output, GigE interface, up to 33 fps, 1/40000 to 1/15 sec shutter speed, C-mount.

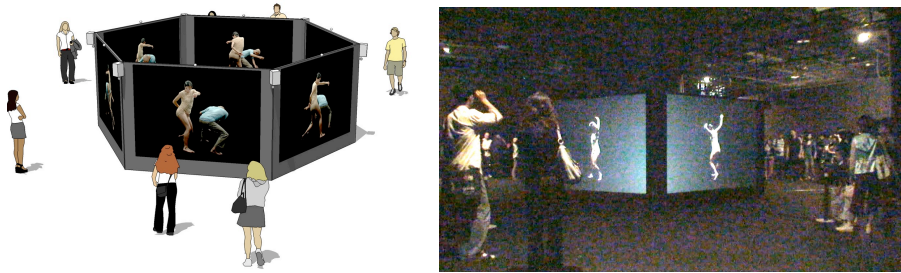


Figure 1,2: Double District in Re-Actor. As a model (l) and at the eArts Festival Shaghai, October 2008 [14] (r)

1.2 Multi-perspective Capture

The modality in which the dance performance was captured, mirrors the physical configuration of the Re-Actor environment. Six evenly distributed stereo camera pairs encircle a 4 m diameter stage. This configuration allows a observer to view the scene from multiple points-of-view, it constitutes multi-perspective capture.

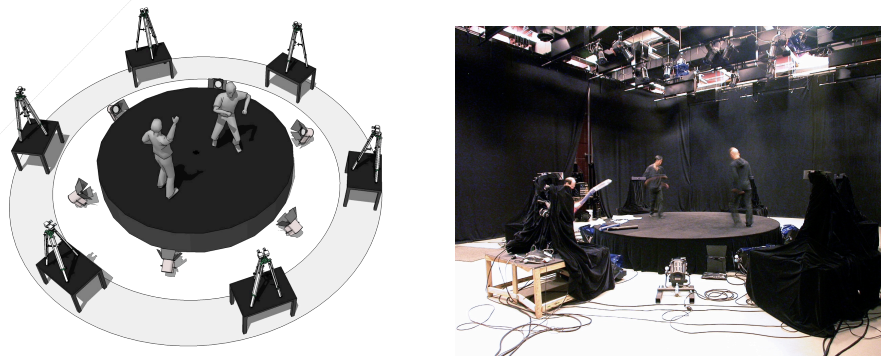


Figure 3,4: Model of recording set-up (l) and in the studio (r)

Precise positioning and orientation of the camera heads is essential to recreate a believable illusion of the physical space on screens. To strengthen the imitation of real world perception on screen, a 10 mm focal length for the camera lenses were chosen to reflect the natural field of view of the human eye [11].



Figure 5: Multi-perspective scene

1.3 Stereographic imaging

The properties of a stereo image capture system are critical for comfortable viewing, inartificial depth perception and the sense of reality a viewer perceives. The relationship between inter-ocular distance, near and far plane, the range of subject movement, focal length and position of the zero parallax plane had to be defined [9]. These parameters were generated in a mathematical model [8] and its values confirmed in an experimental set up. The subjective qualities of the experimental results lead to a minor adjustment of some of the parameters.

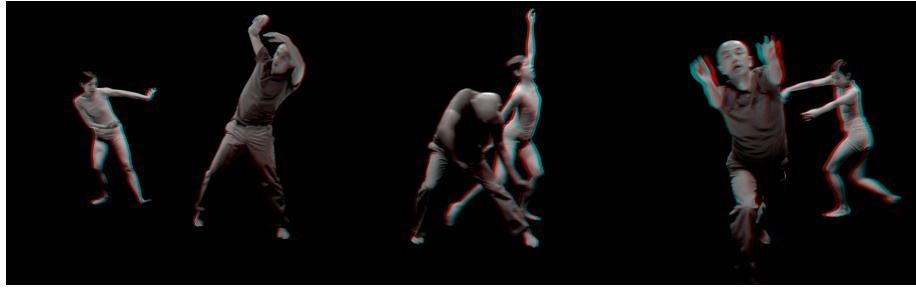


Figure 6: Stereographic video stills in anaglyphic format. The original format is discrete images for the left and right eye.

2. Multi-perspective vs. Universal: Voxelization

This proposed method takes the concept of multi-perspective capture one step further. It uses real-time 3D computer graphic to transform the multi-perspective recording into a universal one [7]. The performance can be observed from any point-of-view, not only from the position of the cameras encircling the scene. The number of cameras does not correlate with the number of possible viewpoints. This is facilitated through volumetric geometry reconstruction of the dance performance, a process named voxelization.



Figure 7,8: A frame of the video in comparison with the same frame and similar perspective for the voxel representation (l). Close-up of a voxel model representing a dancer's torso, head and arms (r).

By geometric calibration of the twelve cameras intrinsic and extrinsic parameters and employing computer vision and image processing algorithms, the parallel and synchronized video streams of the scene are used to synthesize a voxel (Volumetric Pixel) stream [1,2,3,4,5].

Voxels are points in 3D space with a volume attached to them. A larger number of voxels (<5000) defines the geometry accurately enough to be able to recognize elements in the scene and allows for visualisation. In this work, the scene was synthesized with a voxel resolution of ~ 1.5 cm, represented by a cube of this size as the smallest unit. Through averaging color values of the calibrated video stream pixels, RGB color values could be extracted for every voxel.

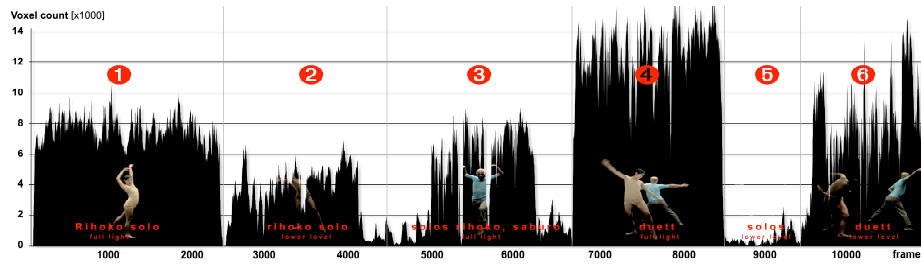


Figure 9: Diagram of the voxel density across time and scene. The y-axis represents the voxel count (x1000) and x-axis the frames in the video.

The number of voxels or their density in the voxel space varies over time and with the complexity of the scene. A solo performance does use a lot less voxels then for instance a duet (Fig.9).

The original studio recordings were not lit to optimize voxel reconstruction, but for artistic and cinematographic reasons alone. Lighting and the less the ideal positioning of the cameras result in a relative low voxel count in some of the scenes, causing a degradation in reproduction quality. For instance a leg is not visible in voxel space due to the fact that it was not lit adequately. A selection process was necessary to pick scenes from the performance with high enough voxel count (<5000 average). In Figure 9, only scenes 1,3 (solos) and 4 (a duet) were kept as bases to work with for the prototype application.

Even then, there are still moments in the performance where the voxel model deteriorates, but this has only limited relevance, the video and its parallel voxel stream do refresh with 30 frames per second and human perception is capable to reconstruct incomplete geometry in motion and make sense of the scene.

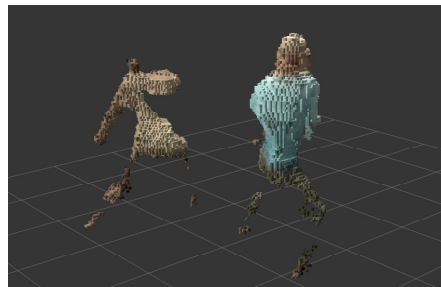


Figure 10: Non adequate lighting and occlusions of the two performers can cause incomplete voxel models.

Ultimately a performance should be captured again, with a similar set-up for the video cameras, but additional multiple infrared cameras, distributed around the stage and pointing down from the ceiling (to avoid occlusions). These cameras together with infrared lighting would produce a much more accurate, in terms of resolution and volume, voxel representation than only the video cameras. Both parallel lighting modes (artistic with theatre lights and infrared) would not interfere with each other due to different wavelengths of the light.

3. Application

An application capable of displaying multiple channels of video (the six multi-perspective video streams) and simultaneously the 3D voxel representation was prototypal developed in Quartz Composer³.

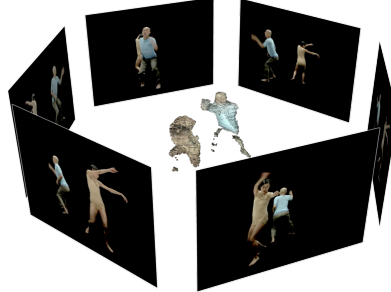


Figure 11: Model of the scene with the six camera views and the voxel representation in the center.

It does allow for navigation in the 3D scene of video and voxel model, keeps track of the synchronicity of the video and the voxel stream and presents time control functions (play, pause, previous/next frame). It snaps the virtual, by the user controlled, camera in place if it gets close to the position of a real video camera, so the perspective of the video image and voxel model corresponds and a seamless fade can be performed.

A list of parameters can be set during runtime: frames per second, point of view, field of view, lighting of the scene and a range of other variables manipulating the aesthetics of the scene and the voxel render style (Fig. 12). The prototype does do all of this in real-time on a MacBook Pro and with a good frame rate, the video resolution is 1024x768 pixel.



Figure 12: Different voxel render styles. Variable size of point-cloud elements.

To be able to take advantage of the full video resolution (1400x1050) and presenting the stereoscopic video and voxel model in stereo/3D (through a passive stereo two projector set-up with polarized filters, glasses and a silver screen) it will be necessary to upgrade to a high performance computer with high-end graphic board and perhaps a different software development environment has to be utilized.

³ Quartz Composer, a node-based visual programming language, part of the Apple Xcode development environment in Mac OS X, based on the Quartz engine, Core Image and OpenGL.

4. Installation and Interaction Modalities

The installation consists of a single stereo-3D projection screen, a console with user interface and two to four sculptures of voxel models made with a rapid prototyping 3D printer (~20 cm high).

The projection shows one of the six camera views full-screen in stereo/3D. The three and a half minute long video segment does run in a loop. With the interface, a visitor is able to change his perspective on the dance scene and is either presented with the real video recording or the voxel representation. The transition between the two modes is seamless due to identical positioning of the virtual and real camera, time synchronicity and equivalent stereo perception parameters.

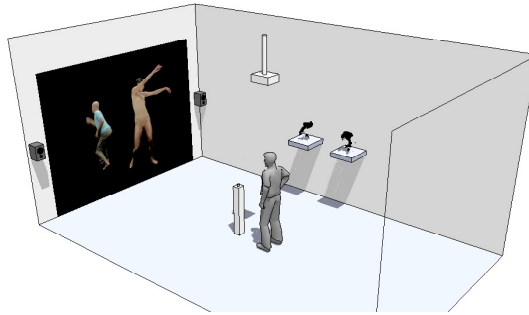


Figure 13: Installation model with passive stereo screen, user interface and voxel sculptures on shelves.

To lessen the time a visitor needs to understand the interaction modalities and his cognitive load, he has only limited freedom to interact with the 3D Video geometry of the scene. A simple rotary controller with push button functionality (Griffin Powermate) allows the user to rotate the scene 360 degree, by doing so the scene will snap into place at the position of a real camera. Using the push button will translate the gaze to a bird's eye view.

5. Conclusion

Multi-perspective recording in combination with voxelization offers a universal view on a scene. A viewer is not limited to one point-of-view or moment in time, but can explore and analyze a scene freely and without space or time restrictions. The event is captured four-dimensional (x, y, depth and time) through the stereoscopic video recording and in post-processing, three additional dimensions are added (x, y, z of the voxel space).

The fidelity and high level of detail in the video imagery is augmented and completed with the voxel representation. Both have different qualities and these are clearly perceived by a viewer, but the fact that the scene is in motion and everything runs in time and space synchronicity helps to get past the gap in visual depiction.

The proposed method constitutes a novel way of recording and documenting motion. It enables detailed analyzation after the event happened. Potential areas of application are in performing arts, in professional sport and the Movie FX industry. This method has the potential to evolve quickly with technological advances. Cameras with higher resolution and depth sensors, better computer vision algorithms and faster processors will eventually be able to create a 3D model with enough detail so video imagery is no longer needed, but for the moment this method delivers universal view today.

Acknowledgements

Double District, 2008

Direction, choreography, lighting design and costumes: Saburo Teshigawara.

Developed with: Volker Kuchelmeister

Performed by: Saburo Teshigawara and Rihoko Sato

Production manager, technical director, stereoscopic cinematography, video and audio post-production: Volker Kuchelmeister (iCinema)

Lighting design: Paul Nichola, Lighting technician: Rob Kelly (NIDA), Production assistant: Sue Midgely (iCinema)

Producer: Richard Castelli (Epidemic)

Co-produced by: Karas (Tokyo), Epidemic (Paris), Le Volcan Scène Nationale (Le Havre), UNSW University of New South Wales iCinema Centre (Sydney) and kindly supported by Museum Victoria.

Voxel reconstruction: Anuraag Sridhar, UNSW School of Computer Science and Engineering

Re-Actor

Re-Actor created by Sarah Kenderdine and Jeffrey Shaw. Originally developed for their virtual 3D theater work UNMAKEABLELOVE [12,15], Kenderdine and Shaw also conceived its use as a 3D visualization architecture for the multi-view presentation of live performances.

References

1. Steven M. Seitz and Charles R. Dyer. 1999. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 35(2): pages 151-173.
2. Kiriakos N. Kutulakos and Steven M. Seitz. 2000. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3): pages 198-218.
3. Greg Slabaugh, Bruce Culbertson, Tom Malzbender and Ron Schafer. 2001. A survey of methods for volumetric scene reconstruction from photographs. *International Workshop on Volume Graphics*.
4. Zhenyu Yang, Bin Yu, Ross Diankov, Wanmin Wu and Ruzena Bajcsy. 2006. Collaborative Dancing in Tele-immersive Environment in *Proc. of ACM Multimedia (MM'06) (Short Paper)*, Santa Barbara, CA.
5. Klara Nahrstedt, Ruzena Bajcsy, Lisa Wymore, Renata Sheppard, Katherine Mezur. 2008. Computation Model of Human Creativity in Dance Choreography. *Association for the Advancement of Artificial Intelligence (AAAI) Spring Syposia*.
6. Double District: Website and Video Documentation, Kuchelmeister Volker. Retrieved June 10, 2009 (http://www.kuchelmeister.net/prj_saburo.html).
7. Universal capture through stereographic multi-perspective recording and scene reconstruction: Website and Video Documentation, Kuchelmeister Volker. Retrieved June 10, 2009 (http://www.kuchelmeister.net/prj_voxel.html).
8. Bourke Paul. Calculating Stereo Pairs. 1999. University of Western Australia. Retrieved June 10, 2009 (<http://local.wasp.uwa.edu.au/~pbourke/miscellaneous/stereographics/stereorender/index.html>).
9. Bourke Paul. 2007. Stereoscopy, Theory And Practice. Workshop at VSMM 2007, 23 September 2007, Brisbane. Retrieved June 10, 2009 (<http://local.wasp.uwa.edu.au/~pbourke/papers/vsmm2007/index.html>).
10. Bourke P. 2008. Stereoscopic Filming - Achieving an accurate sense of depth and scale. University of Western Australia. Retrieved June 10, 2009 (http://local.wasp.uwa.edu.au/~pbourke/miscellaneous/stereographics/stereo_film/).
11. Hunt et al., Chapman and Hall. 1968. *Light, Color and Vision*. Ltd, London: page 49.
12. Kenderdine, S. & Shaw, J. 2009, The relocation of theatre: Making UNMAKEABLELOVE. *Proceedings of Re:Live, Media Art Histories*, Melbourne, November 2009 (forthcoming).
13. Kenderdine S. 2003. This is not a peep show! The Virtual Room at Melbourne Museum (VROOM). ICHIM 2003.

14. Urbanized Landscape, 2008, 'Re-Actor', 'Double District' and 'UNMAKEABLELOVE', Shanghai eArts Festival, Shanghai, September 2008, p. 146, pp. 141-5.
15. UNMAKEABLELOVE. Website. Retrieved June 10, 2009 (<http://unmakeablelove.org/>).