Introducing 3D Cinematography

Rémi Ronfard
INRIA Rhone Alpes

Gabriel Taubin
Brown University

Recent developments in computer vision and computer graphics, especially in such areas as multiple-view geometry and image-based rendering, are making it possible to generate 3D models of dynamic scenes from multiple cameras at video frame rates. We call this process *3D cinematography* since it extends traditional cinematography from 2D (images) to 3D (solid objects that we can render with photorealistic textures from arbitrary viewpoints) at the same frame rate. Many prototypes have demonstrated this capability in various forms and under different names—for example, virtualized reality, free-viewpoint video, and 3D video. Multiple disciplines and research communities have deployed these efforts. Recent advances have clearly shown the promises of 3D cinematography systems, such as allowing real-time, multiple-camera capture, processing, transmission, and rendering of 3D models of real scenes. Yet, the research community must solve many remaining problems before such systems can be transposed from blue screen studios to the real world.

Related research

In the last 10 years, we have seen advances in two main directions. First, we've seen systems with more and more cameras. Carnegie Mellon University's 3D room had 49 cameras. A team at Stanford built a plenoptic camera using 128 video cameras. In 2002, Yang et al. showed a real-time distributed light-field camera made of 64 video cameras. Some of these systems work in real time (for example, 3D television); however, most achieve their effects in postproduction (for example, 3D cinema). Systems that perform plenoptic sampling from a large number of viewpoints still suffer from ghosting artifacts due to depth variations. Therefore, a second direction of research is geometric reconstruction, which is the only way to correct those effects with a finite number of cameras. State-of-the-art systems combine geometric and photometric reconstruction: for example, the unstructured lumigraph, image-based visual hull, or photo hull.

The book edited by Schreer, Kauff, and Sikora is the single best introduction to 3D video technology, covering 3D data representation and processing, reproduction, display, and applications.[1] Recent reference books provide mathematical details—see Hartley and Zisserman, Ma et al., and Nielsen. In addition, Magnor described current methods for reconstructing 3D geometry from videos.

The first workshop on 3D cinematography took place in New York City in June 2006, held jointly with the IEEE Conference on Computer Vision and Pattern Recognition. During the workshop, several papers described camera array technology and reconstruction and interpolation techniques. Naemura, Tago, and Harashima[2] as well as Tanimoto[3] stressed many of the difficulties facing the light-field or ray-space approach. Narayanan and Kanade reviewed work based on geometric reconstruction. Starck and Hilton showed how to animate virtualized actors by combining techniques borrowed from motion capture with 3D cinematography. Extended versions of the papers presented by Narayanan and Kanade and by Starck and Hilton are included as articles in this *IEEE CG&A* issue. Aloimonos stressed some of the difficulties the geometric approach faces, notably in the steps of detecting and interpreting depth discontinuities and putting them in correspondence across views.

In the light of those communications and discussions from the workshop, we now review the most important issues in 3D cinematography.

Volume

One of the main difficulties in all available methods is the limited volume where the action can be recorded. Shape-from-silhouette techniques require at least six to eight cameras to converge on a single scene volume to correctly reconstruct that volume. Ray-space interpolation requires even more cameras. This is limiting 3D cinematography to toy examples with one or two actors remaining close together at the same place all the time.

To virtualize a larger volume, we must allow partial views that don't cover the entire active volume, but together partition the volume. It turns out that this is not a difficult task. Indeed, it's easy to rewrite most existing methods so they can use partial views, as long as the active space is fully covered. But researchers must devote more work to such cases. When the active volume is sparse, with action taking place in small parts of the volume and moving over

 Published by the IEEE Computer Society

time, automatic camera tracking and motion control might become useful.

Camera placement

Choosing an optimal placement for the physical cameras remains a difficult task. Many early papers chose a regular and symmetric arrangement of the cameras, such as in the 3D room. But experience tells us that regular and symmetric camera placements are usually the worst choice, since they only bring redundant information. For instance, in shape-from-silhouette techniques, it would be a mistake to use cameras that face each other and see more or less the same silhouette. A random, irregular camera arrangement will bring more information. But theoretical or even practical guidelines are lacking in this area. Adding more cameras is not always the solution to all 3D cinematography problems. There are many cases when an additional camera should, in fact, be discarded because it's either in a singular configuration with respect to the ongoing action (for instance, the camera might be aligned to the two main actors and confuse one for the other), or it fails to extract the right background or foreground segmentation.

Resolution

There is a paradox with respect to the virtual camera's image resolution. In principle, virtual view synthesis is an interpolation problem and the image quality of the reconstruction should be bounded by the quality of the available views. Yet, in many cases, there is so much redundancy between views that superresolution is achievable. This is especially true when computing virtual views for stereo or multiple-stereo projection. In that case, the human brain fuses the multiple images into a single 3D view whose resolution can be much higher than any of the original views. Combining research in human visual perception and machine vision is needed to understand how 3D cinematography techniques can achieve the same thing.

Spatio-temporal modeling

A particularly exciting area for future development is the use of learning techniques to bootstrap model-based reconstruction. Cheung, Baker, and Kanade developed one such approach for the limited case of a single actor simultaneously seen by at least 16 cameras. The system learned the appearance of the actor's clothes while performing the reconstruction and used this knowledge in subsequent frames. For such systems to scale up to long sequences of performances involving multiple actors, such a capability will become increasingly important.

Virtualized cinematography

Virtual cinematography has become a well-established section of computer graphics.[4] Extending it to natural scenes is not a trivial task. In some cases, the action is scripted, as in a movie or recorded performance. Automatic camera control is suited for those cases, based on the action script. In other cases, the action might be one of a small number of choices, as in a videogame or an online virtual world.

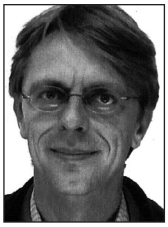
An Internet audience might be satisfied by a system that lets viewers change seats in a virtual theater and focus their attention by panning, tilting, and zooming while viewing the performance. A more ambitious goal is to compute a truly cinematic virtual camera for use by directors and editors in postproduction. Such professional users would not only want to change seats, but to also step on stage, move their camera into the middle of the ongoing action along virtual dolly tracks, and possibly float in the air as if the camera were mounted on a virtual Louma crane. Of course, these types of users would require full-resolution, high-definition, photorealistic views at every step. After 10 years of active research, this technology remains limited to linear arrays of digital cameras along the planned virtual camera path. Recent work has greatly reduced the number of cameras, using a clever combination of multiview stereo blending and inpainting to interpolate among eight cameras along a circular arc. These works have also demonstrated limited extrapolation capabilities.

The next frontier

The next frontier is the synthesis of virtual camera movements[5] along arbitrary paths extrapolating cameras arranged on a plane, a sphere, or even an entire volume. This raises difficult issues. What are the dimensions of the allowable space of cinematographic cameras that professional cinematographers would want to synthesize? In other words, what are the independent parameters of the virtualized cameras that can be interpolated from the set of existing views? Further, what is the range of those parameters that we can achieve using a given physical camera setup? Among those theoretically feasible parameter values, which are the ones that will produce sufficient resolution, photorealism, and subjective image quality? These questions remain open for future research in this new world of 3D cinematography. ■

References

1. O. Schreer, P. Kauff, and T. Sikora, eds., *3D Videocommunication: Algorithms, Concepts, and Real-Time Systems in Human Centred Communication*, Wiley, 2005.
2. T. Naemura, J. Tago, and H. Harashima, "Real-Time Video-Based Modeling and Rendering of 3D Scenes," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, 2002, pp. 66-73.
3. M. Tanimoto, "Overview of Free Viewpoint Television," *Signal Processing: Image Communication*, vol. 21, no. 6, 2006, pp. 454-461.
4. L.-W. He, M.F. Cohen, and D.H. Salesin, "The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing," *Proc. Siggraph*, ACM Press, 1996, pp. 217-224.
5. D. Taylor, "Virtual Camera Movement: The Way of the Future?" *Am. Cinematographer*, vol. 77, no. 9, 1996, pp. 93-100.



Rémi Ronfard *is a senior researcher at INRIA Rhone Alpes. He is also a member of the virtual cinematography research group at ArtificiaLife in Montreal. His research interests include computer vision, computer graphics, and artificial intelligence. Ronfard graduated as a civil engineer and has a PhD in computer science, both from Ecole des Mines de Paris. He was a member of MPEG in 1998 through 2000 and has published 25 papers in the fields of computer vision, computer graphics, and artificial intelligence. Contact him at ronfard@inrialpes.fr.*

Gabriel Taubin *is a professor in the Division of Engineering at Brown University. His research interests include the development of efficient, simple, and mathematically sound algorithms to operate on 3D objects represented as polygonal meshes. Taubin has a Licenciado en Ciencias Matemáticas degree from the University of Buenos Aires, Argentina, and a PhD in electrical engineering from Brown University. He is an IEEE Fellow, winner of the Eurographics 2002 Günter Enderle Best Paper Award, and an IBM Master Inventor. Contact him at taubin@brown.edu.*