Image and Geometry Processing for 3-D Cinematography

Image and Geometry Processing for 3-D Cinematography: An Introduction

Rémi Ronfard and Gabriel Taubin

By 3-D cinematography we refer to techniques to generate 3-D models of dynamic scenes from multiple cameras at video frame rates. Recent developments in computer vision and computer graphics, especially in such areas as multiple-view geometry and image-based rendering have made 3-D cinematography possible. Important applications areas include production of stereoscopic movies, full 3-D animation from multiple videos, special effects for more traditional movies, and broadcasting of multiple-viewpoint television, among others. Drawing from two recent workshop on 3-D Cinematography, the 12 chapters in this book are original contributions by scientists who have contributed to the mathematical foundations of the field and practitioners who have developed working systems.

1 Overview of the Field

The name 3-D cinematography is motivated by the fact that it extends traditional cinematography from 2-D (images) to 3-D (solid objects that can be rendered with photorealistic textures from arbitrary viewpoints) at the same frame rate.

As the name implies, 3-D cinematography focuses on the inter-relations between cinematography (the art of placing cameras and lights to produce good motion pictures) with 3-D modeling. This book summarizes the main contributions from two recent workshops which brought together researchers and practitioners from diverse disciplines, including computer vision, computer graphics, electrical and optical engineering.

R. Ronfard (✉)
INRIA, Grenoble, France,
e-mail: remi.ronfard@inria.fr

G. Taubin
Brown University, Providence, RI, USA,
e-mail: taubin@brown.edu

A first workshop on 3-D Cinematography took place in New York City in June 2006 jointly with the IEEE Conference on Computer Vision and Pattern Recognition.[1] Selected speakers from this workshop were invited to write extended papers, which after review were published as a special section in IEEE Computer Graphics and Applications [7]. At the time some prototypes had demonstrated the ability to reconstruct dynamic 3-D scenes in various forms and resolutions [5, 8]. Various names were used to refer to these systems, such as virtualized reality, free-viewpoint video, and 3-D video. All of these efforts were multi-disciplinary. These advances had clearly shown the promises of 3-D cinematography systems, such as allowing real-time, multiple-camera capture, processing, transmission, and rendering of 3-D models of real dynamic scenes. Yet, many research problems remained to be solved before such systems could be transposed from blue screen studios to the real world.

A second workshop on 3-D Cinematography took place at the BANFF Center in 2008.[2] This workshop focused on summarizing the progress made in the field during the two years subsequent to the first workshop, and in particular on real-time working systems and applications, with an emphasis on recent, realistic models for lights, cameras and actions. Indeed, 3-D cinematography can be regarded as the geometric investigation of lights (how to represent complex lighting, how to relight, etc.); cameras (how to recover the true camera parameters, how to simulate and control virtual cameras, etc.); and actions (how to represent complex movements in a scene, how to edit, etc.).

2   The Geometry of Lights, Cameras and Actions

The theory of 3-D cinematography can be traced back to the discovery of holography by Dennis Gabor in 1946. Gabor firmly believed that holography was destined to replace cinematography as we know it [3]. But to this day, dynamic holograms present overwhelming technological challenges in bandwidth, storage and computing power. Some of today's 3-D cinematography systems follow the more realistic goal of "sampling" the light rays observed from around the scene, as described in Tanimoto's chapter. Most other systems are based on photogrammetry and a partial reconstruction of the 3-D scene. A very good history of those later efforts in 3-D cinematography can be found in the review paper by Kanade and Narayanan [5].

From a geometric viewpoint, it is a hard problem to represent complex, time-varying scenes and their interactions with lights and cameras. One important question explored in this book is – what is the dimensionality of such scenes? Space decomposition methods are popular because they provide approximate answers, although not of very good quality. It has become increasingly evident that better representations are needed. Several partial solutions are proposed in the workshop

---

[1] http://perception.inrialpes.fr/3-Dcine/.

[2] http://www.birs.ca/birspages.php?task=displayevent\&event_id=08w5070.

papers, illustrated with examples. They include wavelet bases, implicit functions defined on a space grid, etc. It appears that a common pattern is the recovery of a controllable model of the scene, such that the resulting images can be edited (interaction).

Changing the viewpoint is only one (important) aspect, but changing the lighting and action is equally important [2]. Recording and representing three-dimensional scenes is an emerging technology made possible by the convergence of optics, geometry and computer science, with many applications in the movie industry, and more generally in entertainment. Note that the invention of cinema (camera and projector) was also primarily a scientific invention that evolved into an art form. We suspect the same thing will probably happen with 3-D movies.

3 Book Contents

The book is composed of 12 chapters, which elaborate on the content of talks given at the BANFF workshop. The chapters are organized into three sections. The first section presents an overview of the inter-relations between the art of cinematography and the science of image and geometry processing; the second section is devoted to recent developments in geometry; and the third section is devoted to recent developments in image processing.

3.1 3-D Cinematography and Applications

The first section of the book presents an overview of the inter-relations between the art of cinematography and the science of image and geometry processing.

The chapter *Stereoscopic Cinema* by Frédéric Devernay and Paul Beardsley is an introduction to stereoscopic 3-D cinematography, focusing on the main sources of visual fatigue which are specific to viewing binocular movies, and on techniques that can be used to produce comfortable 3-D movies. The causes of visual fatigue can be identified and classified into three main categories: geometric differences between both images which cause vertical disparity in some areas of the images, inconsistencies between the 3-D scene being viewed and the proscenium arch (the 3-D screen edges), and discrepancy between the accommodating and the convergence stimuli that are included in the images. For each of these categories, the authors propose solutions to issue warnings during the shooting, or to correct the movies in the post-production phase. These warnings and corrections are made possible by the use of state-of-the-art computer vision algorithms. The chapter explains where to place the two cameras in a real scene to obtain a correct stereoscopic movie. This is what many people understand as 3-D cinematography [4, 6]. In the context of this book, it is worth noting that stereoscopic 3-D recreates the perception of depth from a single viewpoint. In the future, it will be possible for a real

scene to be photographed from a variety of viewpoints, and then its geometry to be fully or partially reconstructed, so that it can later be filmed by two virtual cameras, producing a stereoscopic movie with a variable viewpoint.

One important point made by Devernay and Beardsley is that the success of 3-D cinematography should be measured in terms of *psychological and perceptual* qualities, as can be done for traditional cinema [1]. Their chapter can therefore be a guide for future work in *perceptually-guided* 3-D reconstruction. We can only hope that this leads to a better understanding of depth perception in future immersive and interactive 3-D movies.

The chapter *Free-Viewpoint Television* by Masayuki Tanimoto describes a new type of television, named Free viewpoint Television or FTV, where movies are recorded from a variety of viewpoints. FTV is an innovative visual media that enables users to view 3-D scenes with freedom to interactively change the viewpoint. Geometrically, FTV is based on a ray-space representation, where each ray in real space is assigned a color value. By using this method, Tanimoto and his team constructed the world's first real-time FTV system, which comprises a complete data pipeline, from capturing, to processing and display. They also developed new types of ray capture and display technologies, such as a 360° mirror-scan ray capturing system and a 360° ray-reproducing display. In his chapter, Tanimoto argues that FTV is a natural interface between the viewer and a 3-D environment, and an innovative tool to create new types of content and art.

The chapter *Free-Viewpoint Video for TV Sport Production* by A. Hilton, J.-Y. Guillemaut, J. Kilner, O. Grau and G. Thomas, presents a case study for free-viewpoint television, namely, sports broadcasting. Contrasting Tanimoto's purely ray-based approach, here the authors follow a model-based approach, with geometric models for the field, the players and the ball. More specifically, this chapter reviews the challenges of transferring techniques developed for multiple view reconstruction and free-viewpoint video in a controlled studio environment, to broadcast production for football and rugby. This is illustrated by examples taken from the ongoing development of the *iview* free-viewpoint video system for sports production by the University of Surrey and the BBC. Production requirements and constraints for use of free-viewpoint video technology in live events are identified. Challenges presented by transferring studio technologies to large scale sports stadium are reviewed together with solutions being developed to tackle these problems. This work highlights the need for robust multiple view reconstruction and rendering algorithms which achieve free-viewpoint video, with the quality of broadcast cameras. The advances required for broadcast production also coincide with those of other areas of 3-D cinematography for film and interactive media production.

The chapter *Challenges for Multi-view Video Capture* by Bennett Wilburn further discusses the challenges associated with implementing large scale multi-view video capture systems, with the capture of football matches as a motivating example. This chapter briefly reviews existing multiview video capture architectures, their advantages and disadvantages, and issues in scaling them to large environments. Then it explains that today's viewers are accustomed to a level of realism and resolution which is not feasibly achieved by simply scaling up the performance of existing

systems. The chapter surveys some methods for extending the effective resolution and frame rate of multiview capture systems. It explores the implications of real-time applications for smart camera design and camera array architectures, keeping in mind that real-time performance is a key goal for covering live sporting events. Finally, it comments briefly on some of the remaining challenges for photo-realistic view interpolation of multi-view video for live, unconstrained sporting events.

3.2 Recent Developments in Geometry

The second section of the book presents original contributions to the geometric modeling of large-scale, realistic live-action performances, with an emphasis on the modeling of cameras and actors.

The chapter *Performance Capture from Multi-view Video* by Christian Theobalt, Edilson de Aguiar, Carsten Stoll, Hans-Peter Seidel and Sebastian Thrun, presents an original method to capture performance from a handful of synchronized video streams. The method, which is based on a mesh representation of the human body, captures performance from mesh deformation, and without a kinematic skeleton. In contrast to traditional marker-based capturing methods, this approach does not require optical markings, and it is even able to reconstruct detailed geometry and motion of a dancer wearing a wide skirt. Another important feature of the method is that it reconstructs spatio-temporally coherent geometry, with surface correspondences over time. This is an important prerequisite for post-processing of the captured animations. All of this, the authors argue, can be obtained by capturing a flexible and precise geometrical model of the performers (actors). Their "performance capture" approach has a variety of potential applications in visual media production and the entertainment industry. It enables the creation of high quality 3-D video, a new type of media where the viewer has control over the camera's viewpoint. The captured detailed animations can also be used for visual effects in movies and games.

Most, if not all, 3-D cinematography techniques rely on precise multi-camera calibration methods. The multi-camera calibration problem has been resolved in a laboratory setting, but remains a challenging task in uncontrolled environments such as a theater stage, a sports field, or the set of a live-action movie. The chapter *Combining Multi-view Stereo and Bundle Adjustment for Accurate Camera Calibration* by Yasutaka Furukawa and Jean Ponce, presents a novel approach to camera calibration where top-down information from rough camera parameter estimates and multi-view stereo are used to effectively guide the search for additional image correspondences, and to significantly improve camera calibration parameters using a standard bundle adjustment algorithm.

The chapter *Cell-Based 3-D Video Capture Method with Active Cameras* by Tatsuhisa Yamaguchi, Hiromasa Yoshimoto, and Takashi Matsuyama, deals with the important problem of planning a 3-D cinematographic experiment, in such a way that the best use can be made of a limited number of cameras in a vast area such

as a theater stage or a sports field. 3-D video is usually generated from multi-view videos taken by a group of cameras surrounding an object in action. To generate nice-looking 3-D video, several simultaneous constraints should be satisfied: (1) the cameras should be well calibrated, (2) for each video frame, the 3-D object surface should be well covered by a set of 2D multi-view video frames, and (3) the resolution of the video frames should be high enough to record the object surface texture. From a mathematical point of view, it is almost impossible to find such camera arrangement over a large performance area such as a stadium or a concert stage, where the performers move across the stage. Active motion-controlled cameras can be used in those cases. In this chapter, Matsuyama and co-workers describe a *cellular method* for planning the robotic movements of a set of active cameras, such that the above constraints can be met at all times. Their method is suitable for scripted performances such as music, dance or theater.

The chapter *Dense 3-D Motion Capture from Synchronized Video Streams* by Yasutaka Furukawa and Jean Ponce, describes a novel approach for recovering the deformable motion of a free-form object from synchronized video streams acquired by calibrated cameras. Contrary to most previous work, the instantaneous geometry of the observed scene is represented by a polyhedral mesh with a fixed topology. This represent a very significant step towards applications of 3-D cinematography in computer animation, virtual worlds and video games. The initial mesh is constructed in the first frame using multi-view stereo. Deformable motion is then captured by tracking vertices of the mesh over time, using two optimization processes per frame: a local one using a rigid motion model in the neighborhood of each vertex, and a global one using a regularized nonrigid model for the whole mesh. Qualitative and quantitative experiments using realistic data sets show that this algorithm effectively handles complex nonrigid motions and severe occlusions.

3.3   Recent Developments in Image Processing

The third section of the book presents original contributions in image processing of large-scale, realistic live-action performances, with an emphasis on the modeling, capture and rendering of lighting and texture.

Indirectly estimating light sources from scene images and modeling the light distribution is an important, but difficult problem in computer vision. A practical solution is of value both as input to other computer vision algorithms and in graphics rendering. For instance, photometric stereo and shape from shading requires known light sources. With estimated light such techniques could be applied in everyday environments, outside of controlled laboratory conditions. Light estimated from images is also helpful in augmented reality, to consistently relight artificially introduced objects. Simpler light models use individual point light sources but only work for simple illumination environments. The chapter *Wavelet-Based Inverse Light and Reflectance from Images of a Known Object* by Dana Cobzas, Cameron Upright and Martin Jagersand describes a novel light model using Daubechies wavelets and a

method for recovering light from cast shadows and specular highlights in images. Their model is suitable for complex environments and illuminations. Experiments are presented with both uniform and textured objects and under complex geometry and light conditions. The chapter evaluates the estimation process stability, and the quality of scene relighting. The approach is based on a smooth wavelet representation compared to a non-smooth Haar basis, and on two other popular light representations (a discrete set of infinite light sources and a global spherical harmonics basis).

The chapter *3-D Lighting Environment Estimation with Shading and Shadows* by Takeshi Takai, Susumu Iino, Atsuto Maki, and Takashi Matsuyama, propose the Skeleton Cube to estimate time-varying lighting environments: e.g., lighting by candles and fireworks. A skeleton cube is a hollow cubic object placed in the scene to estimate its surrounding light sources. For the estimation, video of the cube is taken by a calibrated camera and the observed self-shadows and shading patterns are analyzed to compute 3-D distribution of time-varying point light sources. An iterative search algorithm is presented for computing the 3-D light source distribution and several simulation and real world experiments illustrate the effectiveness of the method.

As illustrated in the second section of this book, estimating the full 3-D geometry of a dynamic scene is an essential 3-D cinematography operation for sparse recording setups. When the geometric model and/or the camera calibration are imprecise, however, traditional methods based on multi-view texturing lead to blurring and ghosting artifacts during rendering. The chapter *3-D Cinematography with Approximate or No Geometry* by Martin Eisemann, Timo Stich and Marcus Magnor, presents original image-based strategies to alleviate, and even eliminate, rendering artifacts in the presence of geometry and/or calibration inaccuracies. By keeping the methods general, they can be used in conjunction with many different image-based rendering methods and projective texturing applications.

3-D cinematography of complex live-action scenes with transparencies and multiple levels-of-details requires advances in multi-view image representations. The chapter *View-Dependent Texturing Using a Linear Basis* by Martin Jagersand, Neil Birkbeck and Dana Cobzas, describes a three-scale hierarchical representation of scenes and objects, and explain how it is suitable for both computer vision capture of models from images and efficient photo-realistic graphics rendering. Their model consists of three different scales. The macro-scale is represented by conventional triangulated geometry. The meso-scale is represented as a displacement map. The micro-scale is represented by an appearance basis spanning viewpoint variation in texture space. To demonstrate their model, Jagersand et al. implemented a capture and rendering system based entirely on budget cameras and PC's. For efficient rendering the meso and micro level routines are both coded in graphics hardware using pixel shader code. This maps well to regular consumer PC graphics cards, where capacity for pixel processing is much higher than geometry processing. Thus photo-realistic rendering of complex scenes is possible on mid-grade graphics cards. Their chapter is illustrated with experimental results of capturing and rendering models from regular images of humans and objects.

4   Final Remarks

This book presents an overview of the current research in 3-D cinematography, with an emphasis on the geometry of lights, cameras and actions. Together, the 12 chapters make a convincing point that a clever combination of geometric modeling and image processing is making it possible to capture and render moderately complex dynamic scenes in full 3-D independently of viewpoint.

The next frontier is the synthesis of virtual camera movements along arbitrary paths extrapolating cameras arranged on a plane, a sphere, or even an entire volume. This problem 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 can be achieved using a given physical camera setup? Among the 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 3-D cinematography.

References

1. Cutting, J.: Perceiving scenes in film and in the world. In: J.D. Anderson, B.F. Anderson (eds.) Moving Image Theory: Ecological Considerations, pp. 9–17. University of Southern Illinois Press, Carbondale, IL (2007)
2. Debevec, P.: Virtual cinematography: relighting through computation. Computer **39**(8), 57–65 (2006). doi:http://dx.doi.org/10.1109/MC.2006.285
3. Gabor, D.: Three-dimensional cinema. New Sci. **8**(191), 141–145 (1960)
4. Hummel, R.: 3-D cinematography. American Cinematographer, April (2008)
5. Kanade, T., Narayanan, P.J.: Virtualized reality: perspectives on 4D digitization of dynamic events. IEEE Comput. Graph. Appl. **27**(3), 32–40 (2007). doi:http://dx.doi.org/10.1109/MCG. 2007.72
6. Kozachik, P.: 2 worlds in 3 dimensions. American Cinematographer, February (2009)
7. Ronfard, R., Taubin, G.: Introducing 3D cinematography. IEEE Comput. Graph. Appl. **27**(3), 18–20 (2007). doi:http://dx.doi.org/10.1109/MCG.2007.64
8. Starck, J., Hilton, A.: Surface capture for performance-based animation. IEEE Comput. Graph. Appl. **27**, 21–31 (2007). doi:http://doi.ieeecomputersociety.org/10.1109/MCG.2007.68