Banff International Research Station
for Mathematical Innovation and Discovery

Multi-View Image and Geometry Processing for 3D Cinematography
July 14-18, 2008

MEALS

*Breakfast (Buffet): 7:00–9:30 am, Sally Borden Building, Monday–Friday
*Lunch (Buffet): 11:30 am–1:30 pm, Sally Borden Building, Monday–Friday
*Dinner (Buffet): 5:30–7:30 pm, Sally Borden Building, Sunday–Thursday
Coffee Breaks: As per daily schedule, 2nd floor lounge, Corbett Hall
***Please remember to scan your meal card at the host/hostess station in the dining room for each meal.**

MEETING ROOMS

All lectures will be held in Max Bell 159 (Max Bell Building accessible by walkway on 2nd floor of Corbett Hall). LCD projector, overhead projectors and blackboards are available for presentations. *Please note that the meeting space designated for BIRS is the lower level of Max Bell, Rooms 155–159. Please respect that all other space has been contracted to other Banff Centre guests, including any Food and Beverage in those areas.*

SPONSORSHIP

The workshop is sponsored by Disney Research.

SCHEDULE

Sunday

16:00	Check-in begins (Front Desk - Professional Development Centre - open 24 hours)
17:30–19:30	Buffet Dinner, Sally Borden Building
20:00-22:00	Informal reception in 2nd floor lounge, Corbett Hall Sponsored by Disney Research

Monday

7:00–8:45	Breakfast
8:45–9:00	Introduction and Welcome to BIRS by BIRS Station Manager, Max Bell 159
9:00–10:00	Introduction and Lectures - Lighting
	Introduction (Taubin, Ronfard)
	Raskar, Towards 4D Capture and 6D Displays
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Lectures - Lighting
	Matsuyama, Skeleton Cube: Estimating Time-Varying Lighting Environments
	Cobzas, Smooth and non-smooth wavelet basis for capturing and representing light
11:30–13:30	Lunch
14:00–15:00	Lectures - Cameras
	Furukawa, Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment
	Wilburn, Large scale multiview video capture
15:00–15:30	Coffee Break, 2nd floor lounge, Corbett Hall
15:30–17:30	(Date and time to be confirmed) Guided Tour of The Banff Centre
	meet in the 2nd floor lounge, Corbett Hall
17:30–19:30	Dinner
20:00–20:30	Lectures - Applications
	Beardsley, Capturing Live Action for 3D Cinema

Tuesday

7:00–9:00	Breakfast
9:00–10:00	Lectures - Free-View-Point TV
	Tanimoto, Introducing FTV
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Lectures - Free-View-Point TV
	Aizawa, 3D Video: Generation, Compression and Retrieval
	Tanimoto, Free Viewpoint Audio
11:30–13:30	Lunch
14:00–15:00	Lectures - Stereoscopic Movies
	Zitnick, The filming and editing of stereoscopic movies
	Devernay, Binocular cinematography: 3-D movies for the human eyes
15:00–15:30	Coffee Break, 2nd floor lounge, Corbett Hall
16:00–17:00	Free for open discussions and demos
17:30–19:30	Dinner
20:00–20:30	Lectures - Applications
	Hilton, From 3D Studio Production to Live Sports Events

Wednsday

7:00–9:00	Breakfast
9:00–10:00	Lectures - Action/Motion Capture
	Magnor, Photo-realistic Rendering from Approximate Geometry
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Lectures - Action/Motion Capture
	Theobalt, New Methods for Video-based Performance Capture
	Furukawa, Accurate Camera Calibration from Multi-View Stereo and Bundle
11:30–12:00	Group Photo on the steps of Corbett Hall
12:00–14:00	Picnic and Group Photo on top of hill
14:00–17:30	Free Afternoon
17:30–19:30	Dinner
20:00–20:30	Lectures - Applications
	Ronfard, Automatic Virtual Cinematography

Thursday

7:00–9:00	Breakfast
9:00–10:00	Data Capture / Rendering
	Lanman, New Directions for Active Illumination in 3D Photography
	Jagersand, Hierarchical Model for Capturing and Texturing of 3D Models from 2D Images
10:00–10:30	Coffee Break, 2nd floor lounge, Corbett Hall
10:30–11:30	Data Capture / Rendering
	Matsuyama, Capturing 3D Video of Human Action with a Group of Active Cameras
	Goesele, Multi-View Stereo beyond the Lab Setting
11:30–13:30	Lunch
14:00–15:30	Open Discussions and Closing of workshop
	Grand Challenges and Future Directions
	Closing of workshop
15:30–16:00	Coffee Break, 2nd floor lounge, Corbett Hall
17:30–19:30	Dinner

Friday

7:00–9:00	Breakfast
9:00	Morning is free for informal meetings
10:00	Coffee Break, 2nd floor lounge, Corbett Hall
11:30–13:30	Lunch

Checkout by 12 noon.

** 5-day workshops are welcome to use the BIRS facilities (2nd Floor Lounge, Max Bell Meeting Rooms, Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon. **

Multi-View Image and Geometry Processing for 3D Cinematography
July 14-18, 2008

ABSTRACTS
(in alphabetic order by speaker surname)

Talk 1
Speaker: **Kiyo Aizawa** (University of Tokyo, Japan)
Title: *3D Video: Generation, Compression and Retrieval*
Abstract: We are working on processing of 3D video, which is a sequence of 3D models. 3D video reproduces a real moving object and provides free view point functionality. Differing to CG animation, models in the sequence of 3D video varies in the number of their vertices, connectivities, etc. In our project working together with NHK and ATR in Japan, the quality of capture and reproduction of 3D video is now very much improved. We further work on new topics such as compression, retrieval, editing, etc of 3D video for its wider applications. I would like to talk shortly on generation and focus on compression and retrieval.

Talk 2
Speaker: **Paul Beardsley** (Disney Research Zurich)
Title: *Capturing Live Action for 3D Cinema*
Abstract: Image capture for live-action 3D cinema is traditionally done using a pair of stereo cameras which provide the left-eye and right-eye sequences that will be projected on the cinema screen. This constrains artistic control of 3D effects because decisions about stereo parameters - such as choice of baseline and vergence - are made during shooting, and cannot easily be manipulated afterwards.

This talk describes Disney's new work on a heterogeneous sensor array composed of a cinematographic camera, support cameras, and depth sensors, to shoot live action 3D cinema. The post-production process allows a user to specify a pair of virtual stereo cameras viewing the original scene, with synthetic generation of the left-eye and right-eye images of the virtual rig. Thus stereo parameters, and hence the 3D effects that will be perceived by the viewer in the completed movie, cease to be a fixed and irrevocable choice made when shooting and are instead opened up to artistic control during post-production.

Talk 3
Speaker: **Dana Cobzas** (University of Alberta, Canada)
Title: *Smooth and non-smooth wavelet basis for capturing and representing light*
Abstract: 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. With estimated light such techniques could be applied in everyday environments, outside of controlled lab conditions. Light estimated from images is also helpful in augmented reality in order to consistently relight an artificially introduced object. While algorithms that recover light as individual point light sources work for simple illumination environments, it has been shown that a basis representation achieves better results for complex illumination. In this paper we propose a light model that uses Daubechies wavelets and a method for recovering light from cast shadows and specular highlights in images. We assume that the geometry is known for part of the scene. In everyday images,

one can often obtain a CAD model of man-made objects (e.g. a car), but the rest of the scene is unknown. Experimentally, we tested our method for difficult cases of both uniform and textured objects and under complex geometry and light conditions. We evaluate the stability of estimation and quality of scene relighting using our smooth wavelet representation compared to a non-smooth Haar basis and two other popular light representations (a discrete set of infinite light sources and a global spherical harmonics basis). We show good results using the proposed Daubechies basis on both synthetic and real datasets.

This is joint work with Cameron Upright and Martin Jagersand.

Talk 4
Speaker: **Frederic Devernay** (INRIA, France)
Title: *Binocular cinematography: 3-D movies for the human eyes.*
Abstract: Most often, what is referred to as 3-D movies are really stereoscopic (or binocular) motion images. In stereoscopic motion images, two 2-D movies are displayed, one for the left eye and one for the right eye, and a specific device guarantees that each eye sees only one movie (common devices are active or passive glasses, parallax barrier displays or lenticular displays). 3-D content can be displayed as stereoscopic motion images, but the movie itself does not hold 3-D content, thus the name binocular cinematography. Although shooting a stereoscopic movie seems to be as simple as just adding a second camera, viewing the resulting movie for extended durations can lead to anything from a simple headache to temporary or irreversible damage to the oculomotor function. Although the film industry pushes the wide distribution of 3-D movies, visual fatigue caused by stereoscopic images should still be considered as a safety issue. The main sources of visual fatigue which are specific to viewing binocular movies 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 accommodative and the convergence stimuli that are included in the images. For each of these categories, we propose solutions to either issue warnings during the shooting or correct the movies in the post-production phase. These warning and corrections are made possible by the use of state-of-the-art computer vision algorithms.

Talk 5
Speaker: **Yasutaka Furukawa** (University of Illinois at Urbana-Champaign)
Title: *Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment*
Abstract: The advent of high-resolution digital cameras and sophisticated multi-view stereo algorithms offers the promises of unprecedented geometric fidelity in image-based modeling tasks, but it also puts unprecedented demands on camera calibration to fulfill these promises. In this talk, I will present a novel approach to camera calibration where top-down information from rough camera parameter estimates and the output of a publicly available multi-view-stereo system on scaled-down input images are used to effectively guide the search for additional image correspondences and significantly improve camera calibration parameters using a standard bundle adjustment algorithm. The proposed method has been tested on several real datasets—including objects without salient features for which image correspondences cannot be found in a purely bottom-up fashion, and image-based modeling tasks—including the construction of visual hulls where thin structures are lost without our calibration procedure.

Talk 6
Speaker: **Yasutaka Furukawa** (University of Illinois at Urbana-Champaign)
Title: *Dense 3D Motion Capture from Synchronized Video Streams* Abstract: In this talk, I will propose a novel approach to nonrigid, markerless motion capture from synchronized video streams acquired by calibrated cameras. The instantaneous geometry of the observed scene is represented by a polyhedral mesh with fixed topology. The initial mesh is constructed in the first frame using the publicly available PMVS software for multi-view stereo. Its deformation is captured by tracking its vertices over time, using two optimization processes at each 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 seven real datasets show that our algorithm effectively handles complex nonrigid motions and severe occlusions.

Talk 7
Speaker: **Michael Goesele** (TU Darmstadt, Germany)
Title: Multi-View Stereo beyond the Lab Setting
Abstract: To be announced (Tuesday or Wednsday)

Talk 8
Speaker: **Adrian Hilton** (University of Surrey)
Title: *From 3D Studio Production to Live Sports Events*
Abstract: This talk will review 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. Experience in ongoing development of the iview free-viewpoint video system for sports production in conjunction with the BBC will be presented. Production requirements and constraints for use of free-viewpoint video technology in live events will be identified. Challenges presented by transferring studio technologies to large scale sports stadium will be reviewed together with solutions being developed to tackle these problems. This 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 3D cinematography for film and interactive media production.

Talk 9
Speaker: **Martin Jagersand** (University of Alberta, Canada)
Title: *A Three-tier Hierarchical Model for Capturing and Texturing of 3D Models from 2D Images*
Abstract: We propose a three scale hierarchical representation of scenes and objects and show how this representation is suitable for both computer vision capture of models from images and efficient photo-realistic graphics rendering. The model consists of (1) a conventional triangulated geometry on the macro-scale, (2) a displacement map, introducing pixelwise depth with respect to each planar model facet (triangle) on the meso level. (3) A photo-realistic micro-structure is represented by an appearance basis spanning viewpoint variation in texture space.

To demonstrate the three-tier model we implement a capture and rendering system and show its usefulness on budget cameras and PC's. For capturing the model we use conventional Shape-From-Silhouette for the coarse macro geometry, variational shape and reflectance estimation for the meso-level, and estimate a texture basis for the micro level.

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. We show experimental results capturing and rendering models from regular images of humans and objects.

Talk 10
Speaker: **Douglas Lanman** (Brown University, USA)
Title: *New Directions for Active Illumination in 3D Photography*
In this talk, we focus on recent work on novel 3D capture systems using active illumination by our research group at Brown University. Specifically, we focus on two primary topics: (1) Multi-Flash 3D Photography and (2) Surround Structured Illumination. We conclude by discussing future directions for active illumination in the field, as well as late-breaking work within our laboratory.

Extending the concept of multi-flash photography, we demonstrate how the surface of an object can be reconstructed using the depth discontinuity information captured by a multi-flash camera while the object moves along a known trajectory. Experimental results based on turntable sequences are presented. By

observing the visual motion of depth discontinuities, surface points are accurately reconstructed - including many located deep inside concavities. The method extends well-established differential and global shape-from-silhouette surface reconstruction techniques by incorporating the significant additional information encoded in the depth discontinuities.

We continue our discussion by exploring how planar mirrors can be used to simplify existing structured lighting systems. In particular, we describe a new system for acquiring complete 3D surface models using a single structured light projector, a pair of planar mirrors, and one or more synchronized cameras. We project structured light patterns that illuminate the object from all sides (not just the side of the projector) and are able to observe the object from several vantage points simultaneously. This system requires that projected planes of light be parallel, and so we construct an orthographic projector using a Fresnel lens and a commercial DLP projector. A single Gray code sequence is used to encode a set of vertically-spaced light planes within the scanning volume, and five views of the illuminated object are obtained from a single image of the planar mirrors located behind it. Using each real and virtual camera, we then recover a dense 3D point cloud spanning the entire object surface using traditional structured light algorithms. As we demonstrate, this configuration overcomes a significant hurdle to achieving full 360x360 degree reconstructions using a single structured light sequence by eliminating the need for merging multiple scans or multiplexing several projectors.

Talk 11
Speaker: **Marcus Magnor** (Computer Graphics Lab, TU Braunschweig)
Title: *Photo-realistic Rendering from Approximate Geometry*
Abstract: For 3D cinematography from sparse recording setups, estimating full 3D geometry of the dynamic scene is essential. If the geometry model and/or camera calibration is imprecise, however, multi-view texturing approaches lead to blurring and ghosting artifacts during rendering. In my talk, I will address on-the-fly GPU-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.

Talk 12
Speaker: **Takashi Matsuyama** (Graduate School of Informatics, Kyoto University, Japan)
Title: *Capturing 3D Video of Human Action in a Wide Spread Area with a Group of Active Cameras*
Abstract: 3D video is usually generated from multi-view videos taken by a group of cameras surrounding an object in action. To generate nice-looking 3D video, the following three constraints should be satisfied simultaneously: (1) the cameras should be well calibrated, (2) for each video frame, the 3D 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 enough high to record the object surface texture. From a mathematical point of view, it is almost impossible to find such camera arrangement and/or camera work that satisfy these constraints. Moreover, when an object performs complex actions and/or moves widely, it would be a reasonable way to introduce active cameras to track the object and capture its multi-view videos; otherwise a large number of (fixed) cameras are required to capture video data satisfying the constraints. Then, the fourth constraint is imposed: (4) the group of active cameras should be controled in real time so that each video frame satisfies the above three constraints.

In this talk, we propose what we call a 'Cellular Method' to capture 3D video of human action in a wide spread area with a group of active cameras. In our method, first the problem to find the camera work that satisfies the above four constraints is formulated as an optimization process and then an algorithm to find an optimal solution is presented with experimental results.

This is joint work with H. Yoshimoto, and T. Yamaguchi.

Talk 13
Speaker: **Takashi Matsuyama** (Graduate School of Informatics, Kyoto University, Japan)
Title: *Skeleton Cube: Estimating Time-Varying Lighting Environments*

Abstract: Lighting environments estimation is one of important functions to realize photometric editing of 3D video; lighting can give various effects on 3D video. In this talk, we propose Skeleton Cube to estimate time-varying lighting environmens: e.g. lightings by candles and fireworks. A skeleton cube is a hollow cubic object and located in the scene to estimate its surrounding light sources. For the estimation, video of the cube is taken by a calibrated camera and then observed self shadows and shading are analyzed to compute 3D distribution of time-varying point light sources. We developed an interative search algorithm for computing the 3D light source distribution. Several simulation and real world experiments showed its effectiveness.

This is joint work with T. Takai, and S. Iino.

Talk 14
Speaker: **Ramesh Raskar** (MIT Media Lab)
Title: *Towards 4D Capture and 6D Displays: Mask for Encoding Higher Dimensional Reflectance Fields*
Abstract: In this talk I will describe a capture method that samples 4D reflectance field using a 2D sensor and a display method that encodes 6D reflectance field on 2D film for subsequent viewing.

We capture the 4D reflectance field using a lightfield camera. The lightfield camera used optical spatial heterodyning to multiple sub-aperture views inside a camera.

We describe reversible modulation of 4D light field by inserting a patterned planar mask in the optical path of a lens based camera. We can reconstruct the 4D light field from a 2D camera image without any additional lenses as required by previous light field cameras. The patterned mask attenuates light rays inside the camera instead of bending them, and the attenuation recoverably encodes the ray on the 2D sensor. Our mask-equipped camera focuses just as a traditional camera might to capture conventional 2D photos at full sensor resolution, but the raw pixel values also hold a modulated 4D light field. The light field can be recovered by rearranging the tiles of the 2D Fourier transform of sensor values into 4D planes, and computing the inverse Fourier transform. The lightfield is captured with minimum reduction in resolution allowing a 3D encoding of depth in a traditional photo.

We display 6D reflectance field using a passive mask (2D film) and additional optics. Traditional flat screen displays (bottom left) present 2D images. 3D and 4D displays have been proposed making use of lenslet arrays to shape a fixed outgoing light field for horizontal or bidirectional parallax (top left).

In this article, we present different designs of multi-dimensional displays which passively react to the light of the environment behind. The prototypes physically implement a reflectance field and generate different light fields depending on the incident illumination, for example light falling through a window.

We discretize the incident light field using an optical system, and modulate it with a 2D pattern, creating a flat display which is view *and* illumination-dependent. It is free from electronic components. For distant light and a fixed observer position, we demonstrate a passive optical configuration which directly renders a 4D reflectance field in the real-world illumination behind it. Combining multiple of these devices we build a display that renders a 6D experience, where the incident 2D illumination influences the outgoing light field, both in the spatial and in the angular domain. Possible applications of this technology are time-dependent displays driven by sunlight, object virtualization and programmable light benders / ray blockers without moving parts.

Talk 15
Speaker: **Remi Ronfard** (Xtranormal, Montreal, Canada)
Title: *Automatic Virtual Cinematography*
Abstract: Xtranormal is a Montreal-based company whose mission is to turn movie-making into a universally accessible and social activity. Our research in 3D cinematography is concerned with automating the tasks of placing cameras and lights in a virtual world to create cinematic shots and editing of those shots into a movie. This has applications in *real-time cinematography* for computer games and *scripted cinematography* for movie pre-production. Focusing on the latter case, I will present a quick overview of both traditional and virtual cinematography, including script analysis, shot selection, camera placement and editing, and discuss the issues and opportunities facing this new research area.

Talk 16
Speaker: **Masayuki Tanimoto** (University of Nagoya, Japan)
Title: *Introducing FTV*
Abstract: We have developed a new type of television named FTV (Free viewpoint TV). FTV is an innovative visual media that enables us to view a 3D scene by freely changing our viewpoints. FTV is based on the ray-space method that represents one ray in real space with one point in the ray-space. By using this method, we constructed the world's first real-time FTV system including the complete chain from capturing to display. We have also developed new type of ray capture and display technologies such as a 360-degree mirror-scan ray capturing system and a 360 degree ray-reproducing display. FTV will be widely used since it is an ultimate 3DTV, a natural interface between human and environment, and an innovative tool to create new types of content and art.

Talk 17
Speaker: **Masayuki Tanimoto** (University of Nagoya, Japan)
Title: *FTV with Free Listening-Point Audio*
Abstract: We present novel media integration of 3D audio and visual data for FTV with free listening-point audio. We capture the multi viewpoint and listening-point data, which are completely synchronized, by camera array and microphone array. We have successfully demonstrated the generation of both free viewpoint images and free listening-point audio simultaneously.

Talk 18
Speaker: **Christian Theobalt** (Stanford University, Max Plank Institute)
Title: *New Methods for Video-based Performance Capture*
Abstract: Performance capture means reconstructing models of motion, shape and appearance of a real-world dynamic scene from sensor measurements. To this end, the scene has to be recorded with several cameras or, alternatively, cameras and active scanning devices.

In this talk, I will first present our recent work on mesh-based performance capture from a handful of synchronized video streams. Our method does without a kinematic skeleton and poses performance capture as mesh deformation capture. In contrast to traditional marker-based capturing methods, our approach does not require optical markings and even allows us to reconstruct detailed geometry and motion of a dancer wearing a wide skirt. Another important feature of our method is that it reconstructs spatio-temporally coherent geometry, i.e. we obtain surface correspondences over time. This is an important prerequisite for post-processing of the captured animations.

Performance capture has a variety of potential applications in visual media production and the entertainment industry. As I will show, it enables the creation of high quality 3D 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. I will therefore also briefly talk about ways to post-process the captured data such that they can be modified with off-the-shelf animation software.

Talk 19
Speaker: **Bennett Wilburn** (Microsoft Research Asia)
Title: *Large scale multiview video capture*
Abstract: We discuss issues in large scale multi-view video capture, with football matches as a motivating example. We briefly review existing multiview video capture architectures, their advantages and disadvantages, and issues in scaling them to large environments. We show 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. Thus, we also explore methods for extending the effective resolution and frame rate of multiview capture systems. Real-time performance is a key goal for covering live sporting events, so we explore the implications of real-time applications for smart camera design and camera array architectures. Finally, we comment briefly on some of the remaining challenges for photo-realistic view interpolation of multi-view video for live, unconstrained sporting events.

Talk 20
Speaker: **Larry Zitnick** (Microsoft Research, Redmont, USA)
Title: *The filming and editing of stereoscopic movies*
Abstract: The editing of stereoscopic movies, in which two views are shown to a user to provide the illusion of depth, leads to a variety of novel challenges. For instance when creating cuts between scenes, it is generally desirable to maintain a consistent vergence angle between the eyes. This may be accomplished by careful filming or in post-production using a variety of techniques. In this talk, I will discuss basic video editing tasks in the context of stereoscopic movies, as well as more complex techniques such as the "Hitchcock effect", fade cuts and effects unique to stereoscopic movies.