10411 Abstracts Collection
Computational Video
— Dagstuhl Seminar —

D.Cremers[1], M.A. Magnor[2] and L. Zelnik-Manor[3]

[1] TU München,DE
[2] TU Braunschweig, DE
[3] Technion - Haifa, IL

Abstract. From 10.10.2010 to 15.10.2010, the Dagstuhl Seminar 10411 "Computational Video " was held in Schloss Dagstuhl – Leibniz Center for Informatics. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Video Processing, Image Processing, Computer Vision

10411 Executive Summary – Computational Video

Dagstuhl seminar 10411 "Computational Video" took place October 10-15, 2010. 43 researchers from North America, Asia, and Europe discussed the state-of-the-art, contemporary challenges and future research in imaging, processing, analyzing, modeling, and rendering of real-world, dynamic scenes. The seminar was organized into 11 sessions of presentations, discussions, and special-topic meetings. The seminar brought together junior and senior researchers from computer vision, computer graphics, and image communication, both from academia and industry to address the challenges in computational video. Participants included international experts from Kyoto University, Stanford University, University of British Columbia, University of New Mexico, University of Toronto, MIT, Hebrew University of Jerusalem, Technion - Haifa, ETH Zrich, Heriot-Watt University - Edinburgh, University of Surrey, and University College London as well as professionals from Adobe Systems, BBC Research & Development, Disney Research and Microsoft Research.

Keywords: Video Processing, Image Processing, Computer Vision

Joint work of: Cremers, Daniel; Magnro, marcus A.; Zelnik-Manor, Lihi

Extended Abstract: http://drops.dagstuhl.de/opus/volltexte/2011/2920

Dagstuhl Seminar Proceedings 10411
Computational Video
http://drops.dagstuhl.de/opus/volltexte/2011/2919

High Resolution Passive Facial Performance Capture

Derek Bradley (Disney Research - Zürich, CH)

We introduce a purely passive facial capture approach that uses only an array of video cameras, but requires no template facial geometry, no special makeup or markers, and no active lighting. We obtain initial geometry using multi-view stereo, and then use a novel approach for automatically tracking texture detail across the frames. As a result, we obtain a high-resolution sequence of compatibly triangulated and parameterized meshes. The resulting sequence can be rendered with dynamically captured textures, while also consistently applying texture changes such as virtual makeup.

Convex Relaxation Methods for Computer Vision

Daniel Cremers (TU München, DE)

Numerous computer vision problems can be cast as labelling problems where each point is assigned one of several labels. The case of two labels includes problems like binary segmentation and multiview reconstruction. The case of multiple labels includes problems such as stereo depth reconstruction and image denoising. In my presentation, I will introduce methods of convex relaxation and functional lifting which allow to optimally solve such labelling problems in a spatially continuous setting. Experimental results demonstrate that these spatially continuous approaches provide numerous advantages over spatially discrete (graph cut) formulations, in particular they are easily parallelized (lower runtime), they require less memory (higher resolution) and they do not suffer from metrication errors (better accuracy).

Joint work of: Cremers, Daniel; Kolev, Kalin; Klodt, Maria; Brox, Thomas; Esedoglu, Selim; Pock, Thomas; Chambolle, Antonin

Free-Viewpoint Video with approximate and no geometry

Martin Eisemann (TU Braunschweig, DE)

Free-Viewpoint Video is a new step towards full immersive video, allowing complete control of the viewpoint during playback both in space and time. One major challenge towards this goal is precise scene reconstruction, either implicit or explicit. While some approaches exist which are able to generate a convincing geometry proxy, they are bound to many constraints, e.g., accurate camera calibration and synchronized cameras.

In this presentation I will talk about how to improve rendering quality in a variety of different image-based rendering applications based on our Floating Textures (Eurographics 2008) and about our virtual video camera project

for high quality space-time interpolation (Pacific Graphics 2008, ACM Symposium on Applied Perception in Graphics and Visualization 2008 and Computer Graphics Forum 2010/2011).

Keywords: Image-based rendering, virtual video camera, error concealed rendering

Joint work of: Eisemann, Martin; Stich, Timo; Linz, Christian; Lipski, Christian; Berger, Kai; Sellent, Anita; Rogge, Lorenz; Magnor, Marcus

Deformable Surface Estimation

Peter Eisert (Fraunhofer-Institut - Berlin, DE)

We present methods for the estimation of deformable surfaces from monocular and multi-view video sequences. Different motion models are combined with the optical flow constraint in order to estimate surface deformations in a hierarchical framework. The constant brightness assumption is relaxed by considering changes in illuminations and estimating jointly geometric and photometric properties. Applications of the approach are augmented reality, video stabilization and 3D reconstruction of faces.

A Subspace Approach to Depth of Field Extension in Coded-Aperture Cameras

Paolo Favaro (Heriot-Watt University - Edinburgh, GB)

We present a novel solution to the extension of depth of field in coded-aperture cameras with depth- varying point spread functions (PSF) based on a subspace approach. The proposed solution is based on observing that coded images span well-defined subspaces and, therefore, tools from linear algebra can be used in their analysis. Also, we show how to compare and characterize coded apertures via distances between subspaces and their dimensions. As in previous methods based on imaging systems with a depth-varying PSF, our algorithm recovers both the depth map and the all-in-focus image of the scene from a single input image. However, in contrast to those methods, we show that one can use convolutions with a bank of filters to recover the depth map, rather than deconvolution on test planes. The all-in-focus image can then be recovered by using a deconvolution step with the estimated depth map.

Keywords: Subspace methods, Coded aperture camera, depth of field, blind deconvolution

What's going on? Discovering Spatio-Temporal Dependencies in Dynamic Scenes

Vittorio Ferrari (ETH Zürich, CH)

We present novel models for the analysis of complex dynamic scenes capable of discovering long-term temporal relations between the activities of multiple moving agents (e.g. cars and trams in a traffic scene). The models extend Hierarchical Dirichlet processes to include Hidden Markov Models and can be trained fully automatically directly from a long video of a dynamic scene. Our training procedure infers all parameters such as the optimal number of HMMs necessary to explain a scene. The models discover spatio-temporal rules, such as the right of way between different lanes or typical traffic light sequences. Likely applications are unusual event detection, traffic analysis and video summarization. Results are presented on various real-world traffic scenes from Zurich and London.

Keywords: Dynamic scene analysis; hierarchical dirichlet processes; unsupervised learning

Joint work of: Kuettel, Daniel; Breitenstein, Michael; Van Gool, Luc; Ferrari, Vittorio

Full Paper:
http://www.vision.ee.ethz.ch/∼calvin/publications.html

See also: CVPR 2010

Modelling non-rigid 3D shapes

Andrew Fitzgibbon (Microsoft Research UK - Cambridge, GB)

I will talk about modelling of 3D objects that can change shape, either in video or from sets of photos. I describe two approaches we have used recently: implicit and explicit 3D, and show how they are related. The implicit approach models the image-formation process as a 2D-to-2D transformation directly from an object's texture map to the image, modulated by an object-space occlusion mask, we can recover a representation which we term the "unwrap mosaic". This representation allows a considerable amount of 3D manipulation without ever being explicitly 3D. The second strand is to explicitly recover 3D, for example from a set of photos. Such a set might be obtained by an image search for the term "dolphin". This yields many photos of dolphins, but no two are of exactly the same individual, nor are they the same 3D shape. Yet, to the human observer, this set of images contains enough information to infer the underlying 3D deformable object class. We aim to recover models of such deformable object classes directly from images. For classes where feature-point correspondences can be found, this is a straightforward extension of nonrigid factorization, yielding a

set of 3D basis shapes to explain the 2D data. However, when each image is of a different object instance, surface texture is generally unique to each individual, and does not give rise to usable image point correspondences. We overcome this sparsity using curve correspondences (crease-edge silhouettes or class-specific internal texture edges). Finally I will present some recent work on extending to surface, rather than wireframe, models.

Ambient Point Clouds for View Interpolation

Michael Goesele (TU Darmstadt, DE)

View interpolation and image-based rendering algorithms often produce visual artifacts in regions where the 3D scene geometry is erroneous, uncertain, or incomplete. We introduce ambient point clouds constructed from colored pixels with uncertain depth, which help reduce these artifacts while providing non-photorealistic background coloring and emphasizing reconstructed 3D geometry. Ambient point clouds are created by randomly sampling colored points along the viewing rays associated with uncertain pixels. Our real-time rendering system combines these with more traditional rigid 3D point clouds and colored surface meshes obtained using multi-view stereo. Our resulting system can handle larger-range view transitions with fewer visible artifacts than previous approaches.

Joint work of: Goesele, Michael; Ackermann, Jens; Fuhrmann, Simon; Haubold, Carsten; Klowsky, Ronny ; Steedly, Drew ; Szeliski, Richard

Full Paper:
http://www.gris.informatik.tu-darmstadt.de/~mgoesele/projects/AmbientPointClouds.html

See also: Ambient Point Coulds for View Interpolation Michael Goesele, Jens Ackermann, Simon Fuhrmann, Carsten Haubold, Ronny Klowsky, Drew Steedly, Richard Szeliski In: ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2010), Los Angeles, USA, July 25-29, ACM, New York, 2010.

Convex Relaxations for Multi-Label Problems

Bastian Goldluecke (TU München, DE)

Convex relaxations for continuous multilabel problems have attracted a lot of interest recently.

Unfortunately, in previous methods, the runtime and memory requirements scale linearly in the total number of labels, making them very inefficient and often unapplicable for problems with higher dimensional label spaces. We propose a reduction technique for the case that the label space is a product space, and introduce proper regularizers.

The resulting convex relaxation requires orders of magnitude less memory and computation time than previously, which enables us to apply it to large-scale problems like optic flow, stereo with occlusion detection, and segmentation into a very large number of regions.

Despite the drastic gain in performance, we do not arrive at less accurate solutions than the original relaxation.

Using the novel method, we can for the first time efficiently compute solutions to the optic flow functional which are within provable bounds of typically 5-10% of the global optimum.

Multi-camera capture of live action with broadcast cameras

Oliver Grau (BBC Research & Development - London, GB)

The context of this presentation is the capture of live action using multiple cameras. For this purpose the action - usually human performance - is analysed and a three-dimensional description is extracted using a visual hull computation. This requires camera parameters and segmented label images of the live action and computes a 3D volumetric description of the live action for each set of frames of the camera images.

The main contribution of this presentation reviews the application of this processing pipeline to a real-world production scenario. The characteristics of broadcast cameras are reviewed and their implication to segmentation and camera calibration.

Keywords: Camera sensors, camera calibration, segmentation, camera models, camera respose curve

Stereoscopic Production from Wide-baseline Views

Adrian Hilton (University of Surrey, GB)

Conventional stereoscopic video content production requires use of dedicated stereo camera rigs which is both costly and lacking video editing flexibility. In this paper, we propose a novel approach which only requires a small number of stan- dard cameras sparsely located around a scene to automatically convert the monocular inputs into stereoscopic streams. The approach combines a probabilistic spatio-temporal segmenta- tion framework with a state-of-the-art multi-view graph-cut reconstruction algorithm, thus providing full control of the stereoscopic settings at render time. Results with studio se- quences of complex human motion demonstrate the suitabil- ity of the method for high quality stereoscopic content gener- ation with minimum user interaction.

Keywords: Stereo, 3D Video

A Mathematical Model for Plenoptic Imaging

Ivo Ihrke (Universität des Saarlandes, DE)

I present some recent ideas on a generalized imaging model for future cameras. Instead of just considering a single ray per pixel, I propose to look at the manifold of directional, spectral and temporal variation that is being integrated by a single sensel in our current imaging sensors. While the arrangement of these sensels and the optical systems used could be changed in future designs, the concept of a sensel as an integrating device will most likely persist. This view yields a conceptually simple model of plenoptic imaging devices as projections onto basis functions that every sensel defines. These bases can represent certain subspaces of the plenoptic function.

I present some recent results of analyzing this model for the more restricted setup of sensels laid out on a two-dimensional cartesian grid. I conclude by discussing the promises and challenges of the proposed model.

Keywords: Plenoptic imaging

Colored Illumination to the Rescue

Jan Kautz (University College London, GB)

Colored illumination can be useful in a number of applications, two of which are presented here.

Many vision and graphics problems such as relighting, structured light scanning and photometric stereo, need images of a scene under a number of different illumination conditions. It is typically assumed that the scene is static. To extend such methods to dynamic scenes, dense optical flow can be used to register adjacent frames. This registration becomes inaccurate if the frame rate is too low with respect to the degree of movement in the scenes.

We present a general method that extends time multiplexing with color multiplexing in order to better handle dynamic scenes. Our method allows for packing more illumination information into a single frame, thereby reducing the number of required frames over which optical flow must be computed. Moreover, color-multiplexed frames lend themselves better to reliably computing optical flow. We show that our method produces better results compared to time-multiplexing alone. We demonstrate its application to relighting, structured light scanning and photometric stereo in dynamic scenes.

In the second part of the talk, we will look at flash photography. Flash photography is commonly used in low-light conditions to prevent noise and blurring artifacts. However, flash photography commonly leads to a mismatch between scene illumination and flash illumination, due to the bluish light that flashes emit. Not only does this change the atmosphere of the original scene illumination, it also makes it difficult to perform white balancing because of the illumination differences. Professional photographers sometimes apply colored gel

filters to the flashes in order to match the color temperature. While effective, this is impractical for the casual photographer. We propose a simple but powerful method to automatically match the correlated color temperature of the auxiliary flash light with that of scene illuminations allowing for well-lit photographs while maintaining the atmosphere of the scene. Our technique consists of two main components. We first estimate the correlated color temperature of the scene, e.g., during image preview. We then adjust the color temperature of the flash to the scene's correlated color temperature, which we achieve by placing a small trichromatic LCD in front of the flash. We demonstrate the effectiveness of this approach with a variety of examples.

Keywords: Color multiplexing, time multiplexing, relighting

Full Paper:
http://doclib.uhasselt.be/dspace/handle/1942/10310

See also: CVPR 2009

Extensions to Free-Viewpoint Video with Special Effects and Dynamic Scene Reconstruction

Felix Klose (TU Braunschweig, DE)

The image-based framework of the virtual video camera (Computer Graphics Forum 2010/2011) is capable of creating high quality free-viewpoint video by space-time interpolation. I present how different types of special effects can seamlessly be integrated. (ACM Workshop on 3D Video Processing 2010) Furthermore I show how a patch based scene reconstruction approach can be used on the same input data to simultaneously recover structure and motion of a dynamic scene. (Vision, Modeling and Visualization 2010)

Joint work of: Klose, Felix; Linz, Christian; Lipski, Christian; Berger, Kai; Sellent, Anita; Magnor, Marcus

Effcient Depth-Compensated Interpolation for Full Parallax Displays

Reinhard Koch (Universität Kiel, DE)

Recently new displays have been developed that are able to emit a dense light field granting free viewpoint autostereoscopic vision, which compare to head tracked polarized light stereo displays in quality. These displays require a vast amount of high quality images. Naive rendering of these images directly from the modeling tool would be far too costly and would require months on a standard desktop computer. This paper presents an image based rendering algorithm that is able to render interpolated images from sparse input images considering

the special requirements of the display. Special care is taken to ensure efficient rendering with almost no observable loss in quality and data structures that are able to handle vast amounts of data efficiently.

Keywords: Image-based Rendering, Autostereoscopic full-parallax display

Joint work of: Koch, Reinhard; Jung, Daniel

See also: Daniel Jung and Reinhard Koch: Efficient Depth-Compensated Interpolation for Full Parallax Displays. Proceedings of the Fifth International Symposium on 3D Data Processing, Visualization and Transmission, (3DPVT'10), Paris, France, May 2010.

Time-of-Flight Camera Data Processing and Accumulation

Andreas Kolb (Universität Siegen, DE)

This talk focusses on interactive techniques for processing and accumulation for explorative visualization of time-of-light based imaging sensors, such as the Photonic-Mixing Device (PMD).

The talk will first give a very brief introduction to the Time-of-Flight principle used in current ToF-cameras. Due to various error sources, the PMD range data has to be calibrated and corrected before further usage.

The relative high bandwidth for the delivered range data requires specific approaches for data processing and accumulation. Here, volume and point-based techniques are used to solve these problems.

Keywords: Time-of-Flight, range imaging

Automatic construction of non-rigid 3D scene models from video

Kyros Kutulakos (University of Toronto, CA)

Non-rigidity is pervasive in the world around us: a person's body movements and facial expressions, the deformations of cloth, and the collective motion of a group (e.g., people, cars, plants, etc) can all be described as non-rigid motions in 3D. Unfortunately, capturing non-rigid 3D scene models from a video sequence has proved very hard, and still remains one of the few open problems in visual reconstruction.

In this talk, I will present a new approach to the "non-rigid structure from motion" problem that promises to significantly expand the non-rigid scenes reconstructible from a single video in 3D. The idea is to first solve many local 3-point, N-view *rigid* reconstruction problems independently, providing a "soup" of independently-moving and plausibly-rigid 3D triangles. Triangles in this soup are then combined into deforming bodies in an automatic, bottom-up fashion. I

will show results on a variety of challenging scenes, including deforming cloth, tearing paper, faces, and multiple independently-deforming surfaces.

This is joint work with Jonathan Taylor and Allan Jepson.

Project link: http://www.cs.toronto.edu/~jtaylor/non-rigid/index.html

Full Paper:
http://www.cs.toronto.edu/∼jtaylor/non-rigid/index.html

Tracking and motion classification of swimming microorganisms in 4D digital in-line holographic microscopy data

Laura Leal-Taixe (Universität Hannover, DE)

Digital in-line holography is a microscopy technique which has gotten an increasing amount of attention over the last few years in the fields of microbiology, medicine and physics, as it provides an efficient way of measuring 3D microscopic data over time. In this paper, we present a complete system for the automatic analysis of digital in-line holography data; we detect the 3D positions of the microorganisms, compute their trajectories over time and finally classify these trajectories according to their motion patterns. Tracking is performed using a robust method which evolves from the Hungarian bipartite weighted graph matching algorithm and allows us to deal with newly entering and leaving particles and compensate for missing data and outliers. In order to fully understand the behavior of the microorganisms, we make use of Hidden Markov Models (HMMs) to classify four different motion patterns of a microorganism and to separate multiple patterns occurring within a trajectory. We present a complete set of experiments which show that our tracking method has an accuracy between 76% and 91%, compared to ground truth data. The obtained classification rates on four full sequences (2500 frames) range between 83.5% and 100%.

Keywords: 4D Digital in-line holographic microscopy, Multiple Target Tracking, Motion Classification

Joint work of: Leal-Taixé, Laura; Weiße, Sebastian; Heydt, Matthias; Rosenhahn, Axel; Rosenhahn, Bodo

Full Paper:
http://www.tnt.uni-hannover.de/project/HoloVis/

Efficient Monocular Object Detection and Tracking – Gearing Up for HDTV

Bastian Leibe (RWTH Aachen, DE)

Efficient object detection and tracking is an important component for many video interpretation tasks, ranging from surveillance scenarios and webcam footage to analysis of sports broadcasts. We address the problem of automatically detecting and tracking a variable number of persons in such complex scenes using a monocular, potentially moving camera. We propose an approach for multi-person tracking-by-detection in a particle filtering framework. In addition to final high-confidence detections, our algorithm uses the continuous confidence of pedestrian detectors and online trained, instance-specific classifiers as a graded observation model. Thus, generic object category knowledge is complemented by instance-specific information.

We explore how those unreliable information sources can be used for robust multi-person tracking and demonstrate good tracking performance in a large variety of highly dynamic scenarios.

The move to HDTV brings additional challenges with it. On the one hand side, the additional resolution provided by such footage will make more detailed analysis of sports players feasible in the first place. On the other hand, the increased amount of data will impose even harder constraints on the efficiency of the employed algorithms, in particular of the object detectors. We will analyze the additional processing requirements for current sliding-window detectors in HDTV scenarios and will present an approach how those can be reduced by exploiting dynamic scene geometry constraints.

Online Video Processing

Hendrik P. A. Lensch (Universität Ulm, DE)

The bandwidth and processing power of current GPUs allow for real-time processing of incoming video frames. The tight coupling of video capture and editing allows for a whole set of new applications. I will highlight two scenarios: online temporal filtering and context-aware light projection. In the first project, we aim at providing a camera where the user can tune and program the temporal filtering properties beyond the typical box shaped exposure. Using a high frame rate for the acquisition we can easily accumulate output frames at 60Hz with a user specified weighting function enabling temporal smoothing, sharpening or even calculating Fourier transforms. The second project processes captured video frames and projects out the transformed images with a video projector coaxially aligned with the camera. The processed data is hereby visualized in real-time on the real world object itself rather than on a screen. We can enhance the object's contrast, change its appearance or even accumulate light on the real surface.

Why is Sports Photography Hard? (and what we can do about it)

Marc Levoy (Stanford University, US)

Of all the genres of photography, sports and action is one of the most challenging. Especially for organized team sports played on fields and indoor arenas, the photographer is operating at the limits of pixel sensitivity, burst rate, focusing speed, and lens capabilities offered by modern digital cameras. Of every 1000 pictures the photographer takes, 10-30 are typically usable, depending on experience and luck.

For this challenging problem, computational photography and video offer intriguing possibilities. I will start by describing what's hard about sports photography. I will then enumerate places I think computing can make a difference, especially in focus and tracking, triggering of frames at "decisive moments", and ameliorating the adverse effects of cluttered backgrounds and poor lighting. The talk will be illustrated mainly by my own bad photographs (and a few good ones) of Stanford sporting events.

While this talk may appear tangential to the workshop's theme, it is actually quite relevant. The cameras used by professional sports photographers record full-resolution images at 10 frames per second.

This is getting close to video rate. These cameras don't capture video, but they could, and many other cameras do, although at lower resolution. For each of the challenges in still sports photography, I will consider how video + computation might change the game - improving the photographer's odds and empowering the taking of new kinds of pictures.

Proposal for break-out session: What should be in a programmable video camera?

Marc Levoy (Stanford University, US)

Thesis for discussion: Commercial video cameras are a poor tool for pursuing research in computational video.

First, camcorders do not output raw video. In fact, their video has been so heavily processed that it's hard to know what they are outputting. By contrast, computer vision cameras (like Point Grey) can output raw video, but you couldn't bring one to a soccer game; they have no zoom, focus, stabilization, iris, audio, or built-in power or data storage. Second, camcorders are not programmable. Thus, researchers cannot address tasks that must be done in the camera at the point of capture, like controlling the camera's settings (aperture, shutter, gain, white balance), slewing the camera to track an object (pan, tilt, zoom, focus), knowing when image quality is poor (over/under-exposed, noisy, shaky), etc.

If one were to build an open-source, programmable video camera, what resolution and frame rate should it have? What variables should be controllable? How much memory and computing power should it have? What might we do with it? Is any of this feasible using current technology (without resorting to FPGAs or ASICs)? Bring your complaints about commercial video cameras to this free-for-all discussion. Let's compile a wish-list.

Space Varying Parameter Distributions for Interactive Segmentation

Claudia Nieuwenhuis (TU München, DE)

Segmentation of images or motion fields is often ambiguous without user interaction due to semantic reasons or user intention. Interactive multilable segmentation has, thus, become an important task in computer vision. So far algorithms have focused on color distributions without taking into account the spatial information contained in the user scribbles. We propose to use spatially varying color or model parameter distributions to handle the spatial variability of the objects to segment.

The idea is embedded in a variational minimization problem, which is solved by means of recently proposed convex relaxation techniques. For two regions (i.e. object and background) we obtain globally optimal results for this formulation. For more than two regions the results deviate within very small bounds of about 2 to 4 % from the optimal solution in our experiments. To demonstrate the benefit of spatially variant distributions, we show results for challenging synthetic and real-world examples from the field of image and motion segmentation.

Keywords: Segmentation, interactive, space varying

Joint work of: Nieuwenhuis, Claudia; Cremers, Daniel

3D Shape Reconstruction with Connectivity

Shohei Nobuhara (Kyoto University, JP)

In this talk I will present a robust sparse 3D correspondence estimation algorithm from multi-viewpoint images first, and then introduce a 3D shape reconstruction method which guarantees that the sparse 3D points are included in the 3D shape and connected to each other. With this method we can reconstruct 3D objects with thin and long parts like swords, horns, etc.

Keywords: 3d shape reconstruction

Mid-level representations for videos

Sylvain Paris (Adobe Systems Inc. - Cambridge, US)

We discuss how data representation affects video algorithms. Valuable properties can come for almost free if one carefully chooses an appropriate data structure. We illustrate this idea with extensions of the bilateral grid to video sequences that enable fast and temporal consistent nonlinear processing such as edge-aware smoothing and segmentation, and with the video mesh, a data structure that enables the editing of videos as 3D entities, e.g. to move the camera after the fact.

A first-order primal-dual algorithm for convex problems with applications to imaging

Thomas Pock (TU Graz, AT)

Variational methods have proven to be particularly useful to solve a number of ill-posed inverse imaging problems. In particular variational methods incorporating total variation regularization have become very popular for a number of applications. Unfortunately, these methods are difficult to minimize due to the non-smoothness of the total variation. The aim of this paper is therefore to provide a flexible algorithm which is particularly suitable for non-smooth convex optimization problems in imaging. In particular, we study a first-order primal-dual algorithm for non-smooth convex optimization problems with known saddle-point structure. We prove convergence to a saddle-point with rate O(1/N) in finite dimensions, which is optimal for the complete class of non-smooth problems we are considering in this paper. We further show accelerations of the proposed algorithm to yield optimal rates on easier problems. In particular we show that we can achieve O(1/N 2) convergence on problems, where the primal or the dual objective is uniformly convex, and we can show linear convergence, i.e. O(1/eN) on problems where both are uniformly convex. The wide applicability of the proposed algorithm is demonstrated on several imaging problems such as image denoising, image deconvolution, image inpainting, motion estimation and image segmentation.

A preprint of the paper is available from the webpage www.gpu4vision.org.

Unstructured Video Based Rendering and Modeling

Marc Pollefeys (ETH Zürich, CH)

We present an algorithm designed for navigating around a performance that was filmed as a "casual" multi-view video collection: real-world footage captured on hand held cameras by a few audience members. The objective is to easily navigate in 3D, generating a video-based rendering (VBR) of a performance filmed with widely separated cameras.

Casually filmed events are especially challenging because they yield footage with complicated backgrounds and camera motion. Such challenging conditions preclude the use of most algorithms that depend on correlation-based stereo or 3D shape-from-silhouettes.

Our algorithm builds on the concepts developed for the exploration of photo-collections of empty scenes. Interactive performer-specific view-interpolation is now possible through innovations in interactive rendering and offline-matting relating to i) modeling the foreground subject as video-sprites on billboards, ii) modeling the background geometry with adaptive view-dependent textures, and iii) view interpolation that follows a performer. The billboards are embedded in a simple but realistic reconstruction of the environment. The reconstructed environment provides very effective visual cues for spatial navigation as the user transitions between viewpoints. The prototype is tested on footage from several challenging events, and demonstrates the editorial utility of the whole system and the particular value of our new billboard-to-billboard optimization.

In addition, we also briefly explore the possibility to recover explicit 3D models from hand-held multi-camera datasets.

Keywords: Video-based rendering

Joint work of: Pollefeys, Marc; Ballan, Luca; Puwein, Jens; Taneja, Aparna; Brostow, Gabriel

Full Paper:
http://cvg.ethz.ch/research/unstructured-vbr/

See also: Luca Ballan and Gabriel J. Brostow and Jens Puwein and Marc Pollefeys, Unstructured Video-Based Rendering: Interactive Exploration of Casually Captured Videos, ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010), July, 2010, Los Angeles, pages1–11, ISBN 978-1-4503-0210-4

Non-Chronological Video Manipulations

Yael Pritch (The Hebrew University of Jerusalem, IL)

I will briefly present three topics related to non chronological video manipulations.

(i) In video editing we can manipulations time flow in a way that enables slowing down (or delaying) some dynamic events while speeding up (or advancing) others. Time manipulations are obtained by first constructing an aligned space-time volume from the input video, and then sweeping a continuous 2D slice (time front) through that volume, generating a new sequence of images.

(ii) The same concepts can be used to generate new viewpoints using non-perspective projections (stereo panoramas).

(iii) I will end with Video synopsis, which is the representation of all activities in a long video by a much shorter synopsis video. Synopsis becomes possible by simultaneous display of events that have occurred at different times (non

chronological representation). In addition to video summarization, video synopsis can also serve as an index into the video.

This is joint research with Alex Rav Acha, Moshe Ben Ezra, Dani Lischinski, and Shmuel Peleg

Looking Around Corners: New Opportunities in Ultra-fast Computational Photography

Ramesh Raskar (MIT - Cambridge, US)

Can we photograph objects around corners and beyond the line of sight? This seemingly impossible task can be addressed by analyzing a 5D light transport: 4D + time of flight. Our goal is to exploit the finite speed of light to improve image capture and scene understanding. New theoretical analysis coupled with emerging ultra-high-speed imaging techniques can lead to a new source of computational visual perception. We are developing the theoretical foundation for sensing and reasoning using transient light transport, and experimenting with scenarios in which transient reasoning exposes scene properties that are beyond the reach of traditional machine vision.

Paper: Looking Around the corner using Transient Imaging - Kirmani, Hutchison , Davis, Raskar [in ICCV 2009 Kyoto, Japan]

http://cameraculture.media.mit.edu/femtotransientimaging

Keywords: Time of Flight range imaging, Femto-photography

Full Paper:
http://raskar.info

Full Paper:
http://cameraculture.media.mit.edu/femtotransientimaging

Robust motion fields for alternate exposure and mutli-view video

Anita Sellent (TU Braunschweig, DE)

Most optical flow algorithms consider pairs of images that are acquired with an ideal, short exposure time.

But actually, in times of readily programmable, budget-priced cameras and ample storage space this restriction can be leveraged. We present two approaches, that use additional images of a scene to estimate high accuracy dense correspondence fields.

In our first approach we consider video sequences that are acquired with alternating exposure times so that a short exposed image is followed by a long exposed image that exhibits motion blur. With the help of two enframing short

exposed images, we can decipher not only the motion information encoded in the long exposed image, but also estimate occlusion timings, which are a basis for artifact free frame interpolation.

In our second approach we consider the data modality of multi-view video sequences, as it, e.g., occurs commonly in stereoscopic video. As several images capture nearly the same data of a scene, this redundancy can be used to establish more robust and consistent correspondence fields than the consideration of two images permits.

Keywords: Video correspondence fields

Exploiting the Sparsity of Video Sequences to Efficiently Capture Them

Pradeep Sen (University of New Mexico, US)

Video sequences are known to be highly correlated both spatially and temporally, and this fact is commonly used in transform-domain video compression algorithms such as MPEG. However, until recently it was difficult to see how this fact could be used to accelerate their capture.

In this talk, we describe recent work at the UNM Advanced Graphics Lab on exploiting the sparsity of video sequences to enable their capture using a small number of measurements. Specifically, we leverage the theory of compressed sensing which shows how to reconstruct a signal faithfully from a small number of measurements as long as it is sparse in a transform domain. We present simulated results using a ray tracing system and show that we can get some remarkable video reconstruction by measuring as little as 1

Keywords: Compressed sensing, efficient video capture, video compression

See also: Sen and Darabi "Compressive Rendering" and "Compressive Estimation for Signal Integration in Rendering"

Video Tapestries with Continuous Temporal Zoom

Eli Shechtman (Adobe Systems Inc. - Seattle, US)

I will present a novel approach for summarizing video in the form of a multi-scale image that is continuous in both the spatial domain and across the scale dimension: There are no hard borders between discrete moments in time, and a user can zoom smoothly into the image to reveal additional temporal details. We call these artifacts tapestries because their continuous nature is akin to medieval tapestries and other narrative depictions predating the advent of motion pictures.

We propose a set of criteria for such a summarization, and a series of optimizations motivated by these criteria. These can be performed as an entirely

offline computation to produce high quality renderings, or by adjusting some optimization parameters the later stages can be solved in real time, enabling an interactive interface for video navigation. Our video tapestries combine the best aspects of two common visualizations, providing the visual clarity of DVD chapter menus with the information density and multiple scales of a video editing timeline representation. In addition, they provide continuous transitions between zoom levels.

In a user study, participants preferred both the aesthetics and efficiency of tapestries over other interfaces for visual browsing.

Joint work with Connelly Barnes (Princeton), Dan B Goldman (Adobe) and Adam Finkelstein (Princeton)

Project link: http://www.cs.princeton.edu/gfx/pubs/Barnes_2010_VTW/index.php

Full Paper:
http://www.cs.princeton.edu/gfx/pubs/Barnes_2010_VTW/index.php

See also: Connelly Barnes, Dan B Goldman, Eli Shechtman, and Adam Finkelstein. Video Tapestries with Continuous Temporal Zoom. ACM Transactions on Graphics (Proc. SIGGRAPH) 29(3), August 2010.

Regenerative Morphing

Eli Shechtman (Adobe Systems Inc. - Seattle, US)

In the second part of my talk I will present a new image morphing approach in which the output sequence is regenerated from small pieces of the two source (input) images. The approach does not require manual correspondence, and generates compelling results even when the images are of very different objects (e.g., a cloud and a face). We pose the morphing task as an optimization with the objective of achieving bidirectional similarity of each frame to its neighbors, and also to the source images.

The advantages of this approach are 1) it can operate fully automatically, producing effective results for many sequences (but also supports manual correspondences, when available), 2) ghosting artifacts are minimized, and 3) different parts of the scene move at different rates, yielding more interesting (and less robotic) transitions.

Joint work with: Alex Rav-Acha (Weizmann), Michal Irani (Weizmann), Steve Seitz (UW)

Project webpage: http://grail.cs.washington.edu/projects/regenmorph/

Full Paper:
http://grail.cs.washington.edu/projects/regenmorph/

See also: Shechtman E., Rav-Acha A., Irani M., Seitz S. Regenerative Morphing. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San-Francisco CA, June 2010.

XML3D, XFlow, and AnySL: Making Computational Media and Interactive 3D Available to Everyone via the Web

Philipp Slusallek (Universität des Saarlandes, DE)

The Web has become the main application platform and reaches essentially every computer user – even on mobile devices. With HTML-5 and access to more I/O devices via the DeviceAPI, it even promises to be a strong competitor to the large variety of proprietary App platforms. However, despite the ubiquitous availability of powerful and highly parallel graphics processors even in mobile devices, the Web platform offer only limited interactive 3D graphics capabilities (WebGL) and no access data parallel computing. Making these capabilities available for the Web would allow our research to reach the public directly and thus have a much larger impact.

We are developing three techniques to bring computational media and interactive 3D to the Web: (i) XML3D is a minimal addition to HTML5 that allows to embed interactive 3D graphics directly into any Web page, reusing the existing Web technology wherever possible (including HTML-5, CSS, DOM, events, JavaScript, AJAX, etc.). Interactive HTML and media elements can be applied to 3D geometry; optical, haptic, and physical properties can be assigned and rendered, with minimal and well-known techniques from the Web. Because XML3D objects are just new elements the familiar DOM, any of the millions of web programmer can immediately start programming interactive 3D web applications. (ii) XFlow provides safe data parallel processing in the Web environment by splitting the hard task of developing data parallel kernels from the easy part of applying these kernels to media, 3D, and other data on a Web page. (iii) AnySL finally enables the efficient use of high-level and portable code snippets for different shaders and many small computational kernels by using an embedded compiler to optimize away the cost of a highly modular architecture.

In my talk I will address the possibilities as well as new challenges posed by this new technology and demonstrate it with several examples.

Model-based Editing of 2D and 3D Video

Christian Theobalt (MPI für Informatik - Saarbrücken, DE)

While devices for acquisition and display of 2D video content are ubiquitous, methods and technologies for reconstruction of dynamic 3D content from multiple video streams are still at their infancy. In recent years, however, a variety of techniques where presented that enable us to capture fairly detailed 3D scene representations from multi-view video without fiducial markers in the scene - at least for specific types of scenes, such as moving humans. Many such performance capture techniques rely on a shape model of the scene, which, in the case

of humans, is typically a skeleton model with an attached surface representation or a deformable shape model.

Several performance capture techniques enable the reconstruction of detailed geometry of people in normal everyday attire, including the dynamics and detail of loose fabric, such as a skirt or dress. Despite these recent advances in performance capture technology, it is still hard to conveniently modify such reconstructed 3D scenes. For instance, if one would modify the skeletal motion of the actor after reconstruction, one would not obtain plausible cloth deformation corresponding to the modified motion. I therefore present in this talk a new marker-less method to capture skeletal motion, 3D geometry, and a physics-based simulation model of the character's apparel from multi-view video. With this model, it becomes now feasible to modify a performance after the fact, and to obtain plausible cloth dynamics.

In the second part of the talk, I will show that a model-based marker-less tracking approach also enables previously unseen ways of 2D video editing. From a database of laser scans of humans we learned a parametric model of human shape. This model enables us to modify body by modifying intuitively meaningful parameters, such as height, weight, or muscularity. We can fit the model to a person in a 2D video sequence, and track its motion over time. By modifying the shape parameters of the model and performing image-based warping, we can alter the appearance of the actor according to the modified body parameters. I will show several results in which we spatio-temporally modified the appearance of actors in both 2D and 3D video sequences.

Image-based 3D Modeling via Cheeger Sets

Eno Toeppe (TU München, DE)

We propose a novel variational formulation for generating $3D$ models of objects from a single view. Based on a few user scribbles in an image, the algorithm automatically extracts the object silhouette and subsequently determines a $3D$ volume by minimizing the weighted surface area for a fixed user-specified volume. The respective energy can be efficiently minimized by means of convex relaxation techniques, leading to visually pleasing smooth surfaces within a matter of seconds. In contrast to existing techniques for single-view reconstruction, the proposed method is based on an implicit surface representation and a transparent optimality criterion, assuring high-quality $3D$ models of arbitrary topology with a minimum of user input.

Joint work of: Toeppe, Eno; Oswald, Martin; Cremers, Daniel; Rother, Carsten

3D Video Understanding using a Topology Dictionary

Tony Tung (Kyoto University, JP)

3D video is an imaging technology which consists in a stream of 3D models in motion reconstructed from synchronized multiple view video frames. Each frame is composed of textured 3D models, and therefore the acquisition of long sequences produces massive amounts of data. In order to navigate into datasets and obtain relevant information (such as content description), we propose to model 3D videos using a learned topology-based shape descriptor dictionary.

The dictionary can be either generated from: (1) extracted patterns, or (2) training sequences with semantic annotations, to respectively encode or describe 3D video sequences. The model relies on an unsupervised 3D shape-based clustering of datasets by Reeb graphs and features a Markov network to characterize topology change states in a motion graph.

We show that the use of Reeb graphs as high level topology descriptors is the keystone of the strategy. It allows the dictionary to model complex sequences, especially when the fitting of a 3D skeleton of a priori known topology fails (e.g. subjects wearing loose clothing). The Reeb graph extraction does not require any prior knowledge on the shape and topology of the captured subjects.

Our approach can then achieve content-based compression, skimming and summarization of 3D video sequences using a probabilistic discrimination process.

A semantic description of sequences can be automatically performed as well, thus enabling the system to perform 3D action recognition. Our experiments were carried out on complex 3D videos of real human performances.

Posing to the camera: Automatic viewpoint selection for Human Actions

Lihi Zelnik-Manor (Technion - Haifa, IL)

In many scenarios a scene is filmed by multiple video cameras located at different positions. Viewing these videos simultaneously is hard for the human observer. This raises an immediate question - which camera provides the best view of the scene? Typically, this problem is solved by a human producer that manually selects a single camera to display at each moment in time. Our goal is to automate this process.

In this talk I will first discuss why some viewpoints are "better" than others, and how we define "better". I will then present our first step towards an automatic solution. Currently, given multiple views of the same scene, our method selects those views in which action recognition is easy. We show that the selected views both "look" better intuitively as well as improve automatic action recognition results.

If time permits, I will further discuss future directions and applications of viewpoint selection.

Keywords: Viewpoint selection

How Modern Optic Flow Can Help Computational Video Applications

Henning Zimmer (Universität des Saarlandes, DE)

Many computational video applications such as re-timing or motion capture require a reliable estimation of displacements between subsequent frames.

A promising way to obtain these displacements is to use modern variational optic flow techniques as they have the potential to fulfil the needs of computational video applications.

In this talk we show how to tackle three main requirements:

(i) Robustness under outliers (noise, occlusions, illumination changes),

(ii) an appropriate filling-in of missing information by an appropriate smoothness term and

(iii) an automatic adaptation of the smoothness weight to the given image sequence.

In this context, we show and additionally stress that selecting the appropriate model components for the task under consideration is the key for a favourable performance of the optic flow algorithm.

Keywords: Optic flow, motion estimation, variational methods, robust data term, anisotropic smoothness term, automatic parameter selection

Joint work of: Zimmer, Henning; Bruhn, Andres; Weickert, Joachim; Valgaerts, Levi