Report from Dagstuhl Seminar 11261

Outdoor and Large-Scale Real-World Scene Analysis. 15th Workshop Theoretic Foundations of Computer Vision

Edited by

Frank Dellaert[1], Jan-Michael Frahm[2], Marc Pollefeys[3], and Bodo Rosenhahn[4]

1 **Georgia Institute of Technology, US,** frank@cc.gatech.edu
2 **University of North Carolina – Chapel Hill, US,** jmf@cs.unc.edu
3 **ETH Zürich, CH,** marc.pollefeys@inf.ethz.ch
4 **Leibniz Universität Hannover, DE,** rosenhahn@tnt.uni-hannover.de

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11261 "Outdoor and Large-Scale Real-World Scene Analysis, 15th Workshop Theoretic Foundations of Computer Vision". 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, followed by the scheduled programme.

Overall, the seminar was a great success, which is also reflected in the very positive feedback we received from the evaluation.

Seminar 26. June–01. July, 2011 – www.dagstuhl.de/11261
1998 ACM Subject Classification I.2.6. Learning, I.2.10. Vision
Keywords and phrases Computer Vision, Scene Analysis
Digital Object Identifier 10.4230/DagRep.1.6.54
Edited in cooperation with Laura Leal-Taixé

1 Executive Summary

Frank Dellaert
Jan-Michael Frahm
Marc Pollefeys
Laura Leal-Taixé
Bodo Rosenhahn

The topic of the meeting was *Large-Scale Outdoor Scene Analysis*, which covers all aspects, applications and open problems regarding the performance or design of computer vision algorithms capable of working in outdoor setups and/or large-scale environments. Developing these methods is important for driver assistance, city modeling and reconstruction, virtual tourism, telepresence, and outdoor motion capture. With this meeting we aimed to attain several objectives, outlined below.

A first objective was to take stock of the performance of existing state-of-the-art computer vision algorithms and define metrics and benchmark data-sets on which to evaluate them.

Outdoor and Large-Scale Real-World Scene Analysis. 15th Workshop Theoretic Foundations of Computer Vision, *Dagstuhl Reports*, Vol. 1, Issue 6, pp. 54–80
Editors: Frank Dellaert, Jan-Michael Frahm, Marc Pollefeys, and Bodo Rosenhahn
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

It is imperative that we push existing algorithms, which are currently benchmarked or tested with artificial or indoor set-ups, towards *real* applications. Methods of interest are 3D reconstruction, optic flow computation, motion capture, surveillance, object recognition, and tracking. These need to be dragged out of the lab and into the real world. Over the last years the computer vision community has recognized this problem and several groups are increasingly concentrating on the analysis of uncontrolled scenes. Examples include reconstructing large city models from online image collections such as Flickr, or human tracking and behavior recognition in TV footage or video from arbitrary outdoor scenes. An outcome we envision is the definition of appropriate metrics, benchmark sequences, and the definition of a *grand-challenge problem* that exposes algorithms to all the difficulties associated with large-scale outdoor scenes while simultaneously mobilizing the research community.

A second objective, then, was to define what the open problems are and which aspects of outdoor and large-scale scene analysis make the problem currently intractable. In uncontrolled, outdoor settings many problems start to arise, among them harsh viewing conditions, changing lighting conditions, artifacts from wind, rain, clouds or temperature etc. In addition, large-scale modeling, i.e. spanning city-scale areas, contains difficult challenges of data association and self-consistency that simply do not appear in smaller data-sets. Failure of basic building-block algorithms seems likely or even inevitable, requiring system-level approaches in order to be robust to failure. One of difficulties lies in the fact that the observer looses complete control over the scene, which can become arbitrary complex. This also brings with it the challenge to describe the scene in other than purely geometric terms, i.e., perform true scene *understanding* at multiple spatial and temporal scales. Finally, outdoor scenes are dynamic and changing over time, requiring event learning and understanding as well as integrating behavior recognition. In this, we brought in participants from industry in order to ground the challenges discussed in real-world, useful applications.

The third and final objective was to discuss strategies that address these challenges, by bringing together a diverse set of international researchers with people interested in the applications, e.g. arising from photogrammetry, geoinformatics, driver assistance systems or human motion analysis. Though these people work in different fields and communities, they are unified by their goal of dealing with images and/or video from outdoor scenes and uncontrolled settings. In the workshop we allowed for an exchange of different modeling techniques and experiences researchers have collected. We allowed time for working groups during the workshop that connect people and whose goals are to develop ideas/roadmaps, additionally we allowed young researchers to connect with senior researchers, and in general allow for an exchange between researchers who would usually not meet otherwise.

The seminar schedule was characterised by flexibility, working groups and sufficient time for focused discussions. The participants of this seminar enjoyed the atmosphere and the services at Dagstuhl very much. The quality of this center is unique.

There will be an edited book (within Springer's series on LNCS) following the seminar, and all seminar participants have been invited to contribute with chapters. The deadline for those submissions is in November 2011 (allowing to incorporate results or ideas stimulated by the seminar), and submissions will be reviewed (as normal). Expected publication date is the end of 2012.

2 Table of Contents

Executive Summary

Overview of Talks

Working Groups

3 Overview of Talks

3.1 Bundle Adjustment in the Large

Sameer Agarwal (Google – Seattle, US)

Main reference C. Wu, S. Agarwal, B. Curless, S. Seitz, "Multicore Bundle Adjustment," CVPR 2011.

I will describe the design and implementation of a new Inexact Newton type bundle adjustment algorithm, which uses substantially less time and memory than standard Schur complement based methods, without compromising on the quality of the solution.

Along the way we will revisit the Schur complement trick and see that its use is not limited to factorization-based methods. How it can be used as part of the Conjugate Gradients (CG) method without incurring the computational cost of actually calculating and storing it in memory, and how this is equivalent to the choice of a particular preconditioner. The resulting algorithm is highly parallelizable, and I will describe our multicore CPU and GPU implementations of it.

3.2 Achievements and Challenges in Recognizing and Reconstructing Civil Infrastructure

Ioannis Brilakis (Georgia Institute of Technology, US)

The US National Academy of Engineering has identified restoring and improving urban infrastructure as one of the grand challenges of engineering for the 21st century. Part of this challenge stems from the lack of viable methods to map/label existing infrastructure. For the computer vision community, this challenge becomes "How can we automate the process of extracting geometric, object oriented models of infrastructure from visual data?" Existing methods for object recognition and reconstruction have been successfully adapted to answer this question for small or linear objects (columns, pipes, etc.). However, many civil infrastructure objects are large and/or planar objects without significant and distinctive texture or spatial features, such as walls, doors, windows, floor slabs, and bridge decks. How can we recognize and reconstruct them in a 3D model? In this talk, the speaker will present strategies for infrastructure objects recognition and reconstruction, to set the stage for posing the question above to the audience and initiating the discussion for featureless, large/planar object recognition and modeling.

3.3 I see bad pixels, and they don't even know they're bad!

Gabriel Brostow (University College London, GB)

The limitations of pixel-samples can be viewed from an application-specific perspective. Should this pixel be trusted in the hands of algorithm X? This talk explores how simple supervised learning and computation of everything-we-can-think-of features enables bespoke assessment, measuring the confidence we should have about a pixel's suitability.

Suitability only makes sense when a specific application is defined. To encourage further research into this family of "smart pixels" algorithms, I'll illustrate how we do confidence-assessment for evaluation of i) interest point descriptors, ii) optical flow, iii) Time of Flight , and iv) occlusion regions, as example applications.

3.4 Modeling Temporal Coherence for Optical Flow

Andres Bruhn (Universität des Saarlandes, DE)

Joint work of Volz, Sebastian; Bruhn, Andres; Valgaerts, Levi; Zimmer, Henning
Main reference S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer, "Modeling Temporal Coherence for Optical Flow," Proc. International Conference on Computer Vision (ICCV), 201, accepted for publication

Despite the fact that temporal coherence is undeniably one of the key aspects when processing video data, this concept has hardly been exploited in recent optical flow methods.In this paper, we will present a novel parametrization for multi-frame optical flow computation that naturally enables us to embed the assumption of a temporally coherent spatial flow structure, as well as the assumption that the optical flow is smooth along motion trajectories. While the first assumption is realized by expanding spatial regularization over multiple frames, the second assumption is imposed by two novel first and second order trajectorial smoothness terms. With respect to the latter, we investigate an adaptive decision scheme that makes a local (per pixel) or global (per sequence) selection of the most appropriate model possible. Experiments show the clear superiority of our approach when compared to existing strategies for imposing temporal coherence. Moreover, we demonstrate the state-of-the-art performance of our method by achieving Top 3 results at the widely used Middlebury benchmark.

3.5 Convex Relaxation Techniques for Geometric Optimization

Daniel Cremers (TU München, DE)

I will present recent advances in convex optimization methods for estimating geometry from images. In particular, I will discuss convex formulations of multi-view reconstruction, convex constraints for silhouette consistency and convex formulations for stereo reconstruction. Furthermore I will discuss recent extensions of these optimization techniques to minimal partition problems and to piecewise smooth signal approximation.

3.6 Subgraph Preconditioning: The revolutionary new way of using direct graph-based solvers to speed up conjugate gradient

Frank Dellaert (Georgia Institute of Technology, US)

Direct methods have been very successful in solving large scale, sparse SFM problems. However, when scaling up to graphs with densely connected cliques, the classical “Eiffel-tower" problem, no ordering heuristics can make variable elimination (the basis of all direct methods) fast enough. Based on very recent developments in the theory community, as well as seeing preconditioning as re- parameterization, we now use direct methods to pre-condition the method conjugate gradients. We see this as the way of the future for large-scale, urban structure from motion problems.

3.7 Towards Feature-Based Situation Assessment for Airport Apron Video Surveillance

Ralf Dragon (Leibniz Universität Hannover, DE)

Joint work of Dragon, Ralf;Fenzi, Michele; Shoaib, Muhammad; Rosenhahn, Bodo; Ostermann, Joern
Main reference R. Dragon, M. Shoaib, B. Rosenhahn, and J. Ostermann, “NF-features – no-feature-features for representing non-textured regions,” Proc. ECCV 2010
URL http://www.tnt.uni-hannover.de/papers/view_paper.php?id=842

In this talk, I will give an overview on a project in which we work on a pure feature-based reasoning in an airport apron scenario. Such a medium traffic scenario is hard to assess as background knowledge is crucial (e.g., a car may only pass the runway if no airplane is scheduled). I will explain how a feature-based approach, which is used to extract the current state, is easy to combine with an inference system for large-scale analysis.

I will show, that in feature-based surveillance, the ideas from image-based approaches can be re-used. For example: Foreground or object detection is performed using motion instead of pixel-wise foreground segmentation [1, 2]. Further, methods for feature-based pixel-wise segmentation have been developed [3, 4]. Feature-based object classification can be performed with state-of-the-art object detectors which have high performance for airport apron objects like airplanes, cars or persons [5].

Last but not least, I will discuss the problem of not detecting enough features in a pure feature-based approach. I give an overview on no-feature (NF) features –a feature-based approach to describe non-textured objects– and demonstrate how they improve feature-based background modeling.

References

1 Lauer, Schnörr: Spectral clustering of linear subspaces for motion segmentation, ICCV 2009
2 Toldo, Fusiello: Real-time Incremental J-linkage for Robust Multiple Structures Estimation, ECCV 2010
3 Guillot et al.: Background Subtraction by Keypoint Density Estimation, BMVC 2010
4 Sheikh, Javed, T. Kanade: Background Subtraction for Freely Moving Cameras, ICCV 2009
5 Felzenszwalb, Girshick, McAllester: Object Detection with Discriminatively Trained Part Based Models, TPAMI 32(9), 2010-10

3.8 Homogeneity and inhomogeneity of geometric quality in large scale bundle adjustments

Wolfgang Foerstner (Universität Bonn, DE)

Main reference Richard Steffen, Jan-Michael Frahm and Wolfgang Förstner, “Relative Bundle Adjustment based on Trifocal Constraints,” ECCV 2010.
URL http://www.ipb.uni-bonn.de/uploads/tx_ikgpublication/steffen10_RelativeBundle_RMLE_ECCV2010.pdf

Large scale data acquisition for 3D outdoor scenes requires a homogeneous geometric quality of large areas. This is a severe problem as terrestrial mapping systems need to follow roads, which induce inhomogeneity of the geometric reconstruction as a function of the distance to the acquisition path, a situation known from loop-closing and when including points very far from the sensor path. We discuss means to handle inhomogeneous geometric situations within bundle adjustment (BA), how to specify homogeneity of large BA results using Gaussian processes and to evaluate the geometric quality of BA results.

- http://www.ipb.uni-bonn.de/uploads/tx_ikgpublication/dickscheid08.benchmarking.pdf
- http://www.ipb.uni-bonn.de/uploads/tx_ikgpublication/laebe08.quality.pdf
- http://www.ipb.uni-bonn.de/uploads/tx_ikgpublication/foerstner10_Minimal_Representation_for_Uncertainty-ACCV2010.pdf

3.9 Egomotion estimation and mapping for autonomous systems

Friedrich Fraundorfer (ETH Zürich, CH)

Egomotion estimation and mapping are key tasks for autonomous systems. In this talk I will discuss egomotion estimation and mapping for two examples of autonomous systems, an autonomous car and an autonomous micro aerial vehicle (MAV). In the car example I will discuss egomotion estimation using a monocular camera. I will show how assuming the Ackerman steering model can be used for extremely efficient and robust egomotion estimation and how even absolute scale can be recovered for this monocular case. In the MAV example I will discuss how tight coupling of IMU measurements and visual features lead to extremely efficient and robust egomotion estimation and how the MAV maps its environment for autonomous navigation and obstacle avoidance.

3.10 Objects are More Than Bounding Boxes

Juergen Gall (ETH Zürich, CH)

The goal of object detection is to locate and identify instances of an object category within visual data like images, videos, or 3d data. The location is commonly described by a bounding box and the object categories are based on the human categorization system, e.g., car, bus, pedestrian, etc. For some applications, however, reducing objects to a bounding box and instances of these human categories does not seem to be optimal. In some cases, the task to be solved is more complex. For instance, questions like “Are all cars in this image parking in the right direction?" or “Where can I sit?" cannot be easily answered by classical object detection methods. In this talk, I want to discuss the relevance of object properties for object detection.

3.11 Challenges for Camera-Based Driver Assistance

Stefan Gehrig (Daimler Research – Stuttgart, DE)

One of the many applications of Computer Vision is Camera-Based Driver Assistance. While (almost) every human that owns a driver license is able to perform this task with fault rates of less than 1 accident in 10 years, this simple task becomes extremely challenging for Computer Vision Algorithms. Lanes and objects must be detected and measured at all times, even under adverse weather conditions (rain, snow, backlight, ...). This imposes a high robustness on the algorithms and hence many algorithms developed for controllable environments are deemed inappropriate for such environments. This talk gives an overview of the challenges at hand and shows some directions of how to tackle and solve these challenges. In addition, different ways on how to evaluate the algorithms against‘ ‘weak" ground truth are presented.

3.12 Working with Real-World Data

Michael Goesele (TU Darmstadt, DE)

Joint work of Ackermann, Jens; Curless, Brian; Fuhrmann, Simon; Goesele, Michael; Haubold, Carsten; Hoppe, Hugues; Klowsky, Ronny; Ritz, Martin; Seitz, Steven M., Steedly, Drew; Stork, Andre; Szeliski, Richard

As computer vision researchers, we now enjoy (or fear?) an abundance of real- world data. Well known examples are the billions of images and millions of videos available online. In this talk, I will first present our multi-view stereo and photometric stereo systems able to operate on such real-world data. I will then introduce ambient point clouds as a way to provide a 3D visualization even based on incomplete and uncertain image-based reconstructions of real-world scenes. Finally, I will discuss some challenges, thoughts, and consequences for future work.

3.13 Cross-modal Motion Analysis and Reconstruction

Thomas Helten (MPI für Informatik – Saarbrücken, DE)

Joint work of Tautges, Jochen; Zinke, Arno; Krüger, Björn; Baumann, Jan; Weber, Andreas; Helten, Thomas; Müller, Meinard; Seidel, Hans-Peter; Eberhardt, Bernd
Main reference J. Tautges, A. Zinke, B. Krüger, J. Baumann, A. Weber, T. Helten, M. Müller, H.-P. Seidel, B. Eberhardt, "Motion Reconstruction Using Sparse Accelerometer Data," ACM Transactions on Graphics (May 2011), 30:3, pp.18:1–18:12.
URL http://doi.acm.org/10.1145//1966394.1966397

There are many ways for recording human motion sequences, including optical, inertial and mechanical motion capture (mocap) systems. In particular, optical mocap systems, which provide very rich and easy to interpret data, have been widely used in movie and game productions. However, such systems impose strong restrictions concerning the size of the capture volume and lighting conditions making them difficult to use in outdoor and large-scale real-world scene analysis. Avoiding such restrictions, inertial-based sensors, which have been increasingly used in entertainment and monitoring applications, have become a low-cost alternative for capturing motion characteristics. As a drawback, inertial systems typically deliver rather abstract data such as accelerations and angular velocities, which are prone to noise and difficult to handle. In this contribution, we compare various sensor modalities discussing their strengths and weaknesses. In particular, we address the issue on designing suitable feature representations that allow for cross-modal comparison of motion data. Exemplarily, we illustrate these aspects by means of two application scenarios. Firstly, we describe a method for automatically classifying large-scale trampoline motions on the basis of inertial sensor input. Secondly, we sketch a data-driven approach for reconstructing 3D motions from sparse acceleration data.

3.14 Structure in computer vision and pattern recognition: Why doesn't it really fly high?

Vaclav Hlavac (Czech Technical University, CZ)

Structural pattern recognition has been highly popular in 1960-1970s with seminal contributions from Kung Su Fu. The later interest in it faded with the rise of statistical approaches which penetrated even mathematical linguistics where structural analysis was originally started by Noam Chomsky in 1950s. With images, the failure of structural method was glaring. Recently the structural methods have been mostly understood as the graph embedding (H. Bunke's K.S. Fu Award lecture at ICPR 2010).

In the talk, I will talk about my view of the subject originating from methods from our book Schlesinger M.I., Hlavac V.: "Ten lectures on structural and structural pattern recognition", Kluwer 2002. The structure is also needed for a large scale outdoor scene analysis. I will explain the topic and relate it to my own work and the work of my collaborators: (1) Repetitive structures in house facades, (2) in using structure (pose primitives) in analysis of human activity from video; (3) in using 2D context-free grammars for analysis of mathematical formulae, etc.

3.15 Pyramid Transform Revised: Pyramid Transform on the Manifold

Atsushi Imiya (Chiba University, JP)

Joint work of Imiya, Atsushi; Michizuki, Yoshihiko;Fan, Ming-Ying
Main reference Yoshihiko Mochizuki, Yusuke Kameda, Atsushi Imiya, Tomoya Sakai, Takashi Imaizumi, "Variational method for super-resolution optical flow," Signal Processing, Vol. 91, Issue 7, July 2011, pp. 1535–1567.

The pyramid transform was first proposed in image processing to compress of images preserving global features such as edges and segments.

The transform reduces the size of image data preserving the global features of images by combining shift invariant smoothing and downsampling. Then, the transform provides an efficient strategy for multiresolution image analysis in computer vision.

Multiresolution analysis using pyramid transform allows to unify local and global features on images, which are extracted from low- and high- resolution images, respectively. The pyramid-transform-based method is efficiently used in optical flow computation from planar images captured by pinhole camera systems, since the propagation of features from coarse sampling to fine sampling allows to compute both large-displacement in low-resolution images sampled by a coarse grid and small-displacement in high resolution images sampled by a fine grid. Resizing of an image by downsampling after smoothing by convolution with the Gaussian kernel achieves the image pyramid transform. Since convolution with the Gaussian kernel for smoothing is derived as the solution of diffusion equation, the pyramid transform is achieved by downsampling to the solution of diffusion equation. This separation property of the pyramid transform derives the general pyramid transform on Riemannian manifolds by using downsampling on the manifolds and diffusion equation on manifolds.

As an application, we introduce the Gaussian pyramid transform on the sphere using spherical scale-space analysis and derive a numerically stable optical flow computation algorithm for images on the spherical retina.

There are two typical methods for optical flow estimation for pinhole images; the Lucas-Kanade (LK) method and the Horn-Schunck (HS) method which are based on template matching and the variational minimisation, respectively.Image pyramid technique is commonly used to improve the accuracy of stability of optical flow.For instance, the LK method with pyramid-based multiresolution optical flow computation (LKP) method is derived to guarantee the accuracy and stability of the solution for image sequence using the pyramid transform which detects large-displacement and small-displacement motions. The Gaussian kernel for the pyramid transform is numerically expressed as a discretised small kernel for an planar image, for example, the five times five window is typically selected as the kernel size. For images on the unit sphere, since the grid points of the spherical coordinate are uniformly located,the LKP method is not suitable for optical flow computation on the sphere. However, since variational method such as the HS method only depends on the differentials of a function, the method does not require a uniform grid. Therefore, variational method is suitable for the optical flow computation on the spherical coordinates.

We develop a spherical version of the pyramid-based optical flowcomputation for omni-directional images, since the images captured by any omni-directional imaging system can be transformed to images on the unit sphere.

3.16 Stereo and motion analysis for vision-augmented vehicles

Reinhard Klette (University of Auckland, NZ)

The talk starts with showing current results of stereo and motion analysis for video sequences recorded within a driver assistance project at Auckland university. It points out that data used for performance evaluation do have different properties (e.g. quantified by a SIFT-based measure, see [Haeusler&Klette, Benchmarking Stereo Data – not the Matching Algorithms, DAGM 2010]). Traffic video sequences may be classified into a space of 'situations', defined by combinations of 'events', see[Klette et al, IEEE Trans.

Vehicular Technology 2011]. Trinocular recording [Morales&Klette, Ground truth evaluation of stereo algorithms for real world applications, ACCV workshop, 2010] or approximate modeling of road geometries are options to obtain 'ground truth' for real-world video data.

See www.mi.auckland.ac.nz/EISATS for video test data with various kinds of supporting data towards ground truth.

3.17 Outdoor Ground Truth for Optical Flow?

Daniel Kondermann (Universität Heidelberg, DE)

Creating ground truth for optical flow in natural outdoor environments seems almost impossible. In this talk, I will propose two approaches we are currently investigating.

The first approach is to use semi-automatic vision algorithms as is done for example in movie postproduction to create "pseudo" ground truth. The second approach is to evaluate the properties of today's computer graphic rendering systems with respect to their ability to generate images close to the real world.

Finally I will discuss the problem of defining performance measures and benchmarking with respect to correspondence estimation and related algorithms.

References

1 S. Meister and D. Kondermann. Real versus realistically rendered scenes for optical flow evaluation. In *Electronic Media Technology (CEMT), 2011 14th ITG Conference on*, 2011.

3.18 Two disturbing Remarks on Visual Representations: Hierarchies and Semantics

Norbert Krueger (University of Southern Denmark – Odense, DK)

The advantages of hierarchies with explicit semantics in human and computer vision are discussed. The usage of such hierarchies in today's computer vision is reflected about and a concrete hierarchical system is briefly presented.

3.19 Microscopic vs. Macroscopic crowd analysis

Laura Leal-Taixé (Leibniz Universität Hannover, DE)

Methods for the analysis of crowd videos are usually divided into microscopic and macroscopic. Microscopic methods deal with semi-crowded videos, and the goal is mainly to track each individual over time. Most of these methods are divided into object detection and track linking and occlusions and false alarms are usually the main concerns. On the other hand, what happens when the crowd grows bigger and individuals can no longer be detected? Heavily crowded scenes are handled by macroscopic methods, which have a different goal: finding the general flow of the crowd, preferred paths in a scene, entrances and exits and even panic analysis.

The question that arises is: Can we mix both methods in order to obtain more robust trackers?

3.20 From Image Orientation to Buildings and Trees

Helmut Mayer (Universität der Bundeswehr – München, DE)

The basis of our work is the orientation of images taken from the ground or from small (around 1 kg) unmanned aerial systems (UAS). We employ SIFT points (Lowe 2004), but find correspondences via cross-correlation and least squares matching to obtain highly precise points also for wide baselines and different scales.

Corresponding points are input to Nister's (2004) five point algorithm embedded into a expectation maximization (EM) based robust highly accurate orientation procedure (Bartelsen and Mayer 2010). The procedure can also integrate GPS information for absolute orientation.

Orientation is the basis for object extraction. Our approach for 3D facade interpretation (Reznik and Mayer 2008) based on implicit shape models (Leibe et al. 2004) finds rows or columns of 3D windows. Generative stochastic modeling is at the core of our approach to extract unfoliaged trees from terrestrial images (Huang and Mayer 2009).

3.21 Biologically Inspired Spatiotemporal Saliency Processing for a Computational Model of Visual Attention

Baerbel Mertsching (Universität Paderborn, DE)

In studies of attention, the bottom-up conspicuity of a visual feature is known as *saliency*, describing the level of difference of a feature compared to its spatial neighbors regarding a certain dimension. Common dimensions are color, orientation or size. Traditional, computational models generate saliency maps for these dimensions and combine them into one master map of saliency from which spatial positions with high bottom-up conspicuity can be extracted. Watanabe and Shimojo have shown that in auditory attention saliency exists with temporal aspects: In their setup a single sound can help solve a visual ambiguity while it loses this ability when embedded in an temporal sequence of similar sounds, losing its saliency. In modeling attention, temporal dynamics appear in mechanisms like inhibition of return, which keeps the system from analyzing the same part of the scene again. Mechanisms like that can bee seen as top-down interaction within the model. This work proposes that temporal aspects could already start at the bottom of the system, at the point of saliency processing.

Temporal saliency in vision could work similar than the previously described temporal auditory saliency. It is observable in biological vision, that temporal events like the onset of a stimulus attract attention. Also, it is plausible, that the same onset loses this ability when embedded in a temporal sequence of similar onsets.

3.22 Towards an integrated approach to motion analysis and segmentation

Rudolf Mester (Universität Frankfurt, DE)

What is needed today for advancing further towards practically useful systems for dynamic scene analysis is essentially twofold:

On one hand, the rich repertoire of available techniques for tasks such as optical flow estimation, stereo, segmentation etc., has to be reviewed from the point of view of theoretical sustainability: are all the criteria used in current algorithms solidly based on physical reality, and do they consider 'quality' and 'reliability' as a statistical concept? This necessarily leads to a statistical and signal-theoretic derivation of the optimization target functions ('energy functions') which characterize contemporary algorithms.

It implies also that we need to model images and sequences as the result of a compound process of objects (or regions) where the processes of object (or region) generation, their motion, and the mapping from scene to image are described in a physically realistic way.

The result is a rich, generative 'forward model' which is conceptually built on likelihoods and model parameters learnt from reality, and less on ad hoc energies.

Secondly, the inverse process of inferring from observed images onto the real composition and evolution of the scene, should be addressed in a recursive, Bayesian-filter-like manner, where at each time instant the complete state of information obtained in the past is fused with current observations.

3.23 Realtime 3D Motion Reconstruction from Depth Camera Input

Meinard Mueller (MPI für Informatik – Saarbrücken, DE)

Joint work of Baak, Andreas; Müller, Meinard; Theobalt, Christian; Seidel, Hans-Peter

The reconstruction of human motions from sensor input constitutes a challenging problem in computer vision with numerous applications in biomechanics, medicine, and computer animation. While marker-based approaches constitute a reliable and well understood technique to obtain high-quality motions, they require a significant amount of expensive hardware and intrusive equipment to be attached to the actor's body. On the other side, markerless human pose estimation from multiple video streams becomes extremely difficult when using only few cameras or when tracking outdoor scenes. Recently, depth cameras such as the Microsoft Kinect have shown great potential for obtaining reasonable 3D pose estimates even from a single depth image stream. In this contribution, we describe a method that allows for tracking full-body human motions from a single depth image stream captured in natural, non-intrusive settings. We present a hybrid strategy where the tracking is driven by local optimization component and stabilized by a global data-driven retrieval component. Our experiments show that one obtains stable pose estimation results even for fast and complex motions at real-time frame rates.

3.24 Outdoor Human Motion Capture with Data-Driven Manifold Sampling

Gerard Pons-Moll (Leibniz Universität Hannover, DE)

Joint work of Pons-Moll, Gerard; Baak, Andreas; Gall, Juergen; Mueller, Meinard; Seidel, Hans-Peter; Rosenhahn, Bodo

Main reference Pons-Moll G., Baak A., Gall J., Leal-Taixé L., Mueller M., Seidel H.P, Rosenhahn B, "Outdoor Human Motion Capture with Data-Driven Manifold Sampling," International Conference on Computer Vision, 2011

Human motion capturing (HMC) from multiview image sequences constitutes an extremely difficult problem due to depth and orientation ambiguities and the high dimensionality of the state space. In this paper, we introduce a novel hybrid HMC system that combines video input with sparse inertial sensor input.

Employing an annealing particle-based optimization scheme, our idea is to use orientation cues derived from the inertial input to sample particles from the manifold of valid poses. Then, visual cues derived from the video input are used to weight these particles and to iteratively derive the final pose. As our main contribution, we propose an efficient sampling procedure where hypothesis are derived analytically using state decomposition and inverse kinematics on the orientation cues. Additionally, we introduce a novel sensor noise model to account for uncertainties based on the von Mises-Fisher distribution. Doing so, orientation constraints are naturally fulfilled and the number of needed particles can be kept very small. More generally, our method can be used to sample poses that fulfill arbitrary orientation or positional kinematic constraints. In the experiments, we show that our system can track even highly dynamic motions in an outdoor setting with changing illumination, background clutter, and shadows.

3.25 Affine-invariant diffusion geometry for the analysis of deformable 3D shapes

Dan Raviv (Technion – Haifa, IL)

Joint work of Raviv, Dan; Bronstein, Alexander; Bronstein, Michael; Kimmel, Ron; Sochen, Nir
Main reference IEEE Computer Vision and Pattern Recognition (CVPR) 2011

We introduce an (equi-)affine invariant diffusion geometry by which surfaces that go through squeeze and shear transformations can still be properly analyzed.

The definition of an affine invariant metric enables us to construct an invariant Laplacian from which local and global geometric structures are extracted.

Applications of the proposed framework demonstrate its power in generalizing and enriching the existing set of tools for shape analysis.

3.26 Large Scale Traffic Scene Analysis with Multiple Camera Systems

Ralf Reulke (HU Berlin, DE)

The use of camera systems for traffic monitoring is obvious and already in use for about 20 years. However, most of the cameras are operated locally (similar to an induction loop). It has been shown that the two-dimensional areal data analysis offers new possibilities for the determination of traffic parameters.

These presentations discuss a trajectory based recognition algorithm for atypical event detection in multi object traffic scenes and to obtain area based types of information (e.g. maps of speed patterns, trajectory curvatures or erratic movements). Different views of the same area by more than one camera are necessary, because of the typical limitations of single camera systems, resulting from occlusions by other cars, trees and traffic signs. Furthermore, distributed cooperative multi-camera system (MCS) enables a significant enlargement of the observation area. The fusion of object data from different cameras is done by a multi-target tracking approach. This approach opens up opportunities to identify and specify traffic objects, their location, speed and other characteristic object information. The use of wide baseline stereo methods can improve object detection and the tracking accuracy. An approach, which describes the interaction of traffic objects, will also be presented.

3.27 Towards Fast Image-Based Localization on a City-Scale

Torsten Sattler (RWTH Aachen, DE)

Joint work of Sattler, Torsten; Leibe, Bastian; Kobbelt, Leif
Main reference T. Sattler and B. Leibe and L. Kobbelt, “Fast Image-Based Localization using Direct 2D-to-3D Matching,” ICCV 2011 (to appear).

Image-based localization via pose estimation constitutes an important step in many interesting Computer Vision applications such as tourist navigation, augmented reality and incremental Structure-from-Motion. With the advent of large scale reconstructions, computing correspondences between 2D features and 3D points in the model quickly becomes the main bottleneck in the pose estimation pipeline. Current state-of-the-art localization methods thus try to leverage the structure of the models in order to limit the search space. In this talk we present a simple method that directly established 2D-to-3D correspondences without needing to exploit those structures. Using a prioritization scheme based on visual words allows our method to efficiently handle large, (nearly) city-scale models while outperforming the more complex current state-of-the-art methods both in terms of speed and accuracy.

3.28 Semantic Structure from Motion

Silvio Savarese (University of Michigan, US)

We propose a new framework called Semantic Structure from Motion (SSFM) for jointly recognizing objects as well as reconstructing the underlying 3D geometry of the scene (cameras, points and objects). In our SSFM framework we leverage the intuition that measurements of keypoints and objects must be semantically and geometrically consistent across view points. Our framework has the ability to: i) estimate camera poses from object detections only; ii) enhance camera pose estimation, compared to feature-point-based SFM algorithms; iii) improve object detections given multiple uncalibrated images, compared to independently detecting objects in single images. Extensive quantitative results on three datasets ‘LiDAR cars, street-view pedestrians, and Kinect office desktop" verify our theoretical claims.

3.29 Measuring and Modeling the World – Bayes and Analysis by Synthesis

Andreas Schilling (Universität Tübingen, DE)

Computer Vision deals with modeling from images. This goal can be defined as finding the most probable of all models that could have produced the measured data, i.e. the taken images. A generative approach to reaching this goal consists in minimizing the distance between the input images and images rendered from the model. This is the classical analysis-by-synthesis approach which can be considered optimal in the sense that this image distance is inversely related to the likelihood of the data.

Bayesian reconstruction tries to maximize the posterior probability of the model which is the product of the likelihood and a prior probability of the model. An impressive example for the power of the analysis-by-synthesis approach is the reconstruction of textures from several obliquely taken images. Important open questions concerning analysis-by-synthesis methods include: 1.) finding good models and representations, that allow for good regularization techniques in the case of ill posed vision problems (bringing in prior information about the model probability), 2.) efficient optimization techniques and initializations, as well as 3.) finding filters or transformations, that remove information from the images to be compared that is a consequence of effects not modeled by the class of models under consideration.

3.30 Segmentation, Classification and Reconstruction of Surfaces from Point Clouds of Man-made Objects

Falko Schindler (Universität Bonn, DE)

Joint work of Schindler, Falko; Förstner, Wolfgang; Frahm, Jan-Michael
Main reference Falko Schindler and Wolfgang Förrstner, "Fast Marching for Robust Surface Segmentation," to appear in: Lecture Notes in Computer Science, 2011.

We present a surface model and reconstruction method for man-made environments taking into account prior knowledge about topology and geometry. The model favors but is not limited to pairwise orthogonal vertical and horizontal planes. We do not require one particular class of sensors, as long as a triangulated point cloud is available. The reconstruction method delivers a complete 3D segmentation, parametrization and classification for surface regions and their inter-plane relations. Starting with a curvature adaptive pre- segmentation we reduce the computational cost and are more robust to noise and outliers. All reasoning is statistically motivated, based on only a few decision variables. We demonstrate our reconstruction method for multi-view stereo and structured light reconstructions as well as for laser range data.

3.31 Non-Perspective Camera Models in Underwater Imaging – Overview and Error Analysis

Anne Sedlazeck (Christian-Albrechts-Universität, Kiel, DE)

Joint work of Sedlazeck, Anne; Koch, Reinhard

When capturing images underwater, image formation is affected in two major ways. First, the light rays traveling underwater are absorbed and scattered depending on their wavelength, creating effects on the image colors. Secondly, the glass interface between air and water refracts the ray entering the camera housing because of a different index of refraction of water, hence the ray is also affected in a geometrical way. This paper examines different camera models and their capabilities to deal with geometrical effects caused by refraction. Using imprecise camera models leads to systematic errors when computing 3D reconstructions or otherwise exploiting geometrical properties of images. In the literature, many authors have published work on underwater imaging by using the perspective pinhole camera model (single viewpoint model – SVP) with a different effective focal length and distortion to compensate for the error induced by refraction at the camera housing. On the other hand, methods were proposed, where refraction is modeled explicitly or where generic, non-single-view-point camera models are used. In addition to discussing all three model categories, an accuracy analysis of using the perspective model on underwater images is given and shows that the perspective model leads to systematic errors that compromise measurement accuracy.

3.32 Outdoor Image-based Motion Capture and Reshaping of Humans

Thorsten Thormaehlen (MPI für Informatik – Saarbrücken, DE)

Joint work of Thormaehlen, Thorsten;Rosenhahn Bodo; Halser Nils; Wand Michael; Gall Juergen; Jain, Arjun, Kurz, Christian; Theobalt, Christian; Seidel, Hans-Peter
Main reference N. Hasler, B. Rosenhahn, T. Thormaehlen, M. Wand, J. Gall, H.-P. Seidel, "Markerless Motion Capture with Unsynchronized Moving Cameras," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami Beach, Florida, June 2009.
URL http://www.mpi-inf.mpg.de/resources/MovieReshape/

In this talk I will present techniques to capture the motion of human subjects, which are recorded with multiple unsynchronized moving cameras in an outdoor environment. If multiple moving cameras record the same scene, a camera is often visible in another camera's field of view. This poses a constraint on the position of the observed camera, which can be included into the camera motion optimization process. In cluttered outdoor scenes, silhouettes for human motion estimation are difficult to obtain. We show that reliable estimates are nevertheless possible, if the parameters of the background segmentation are simultaneously updated. Once the camera motion and motion of a human subject has been established, semantically meaningful attributes of body shape, such as height, weight, or waist girth, can be interactively modified. This enables spatio-temporal reshaping of human subjects in outdoor video.

3.33 Inverse Procedural Modeling

Michael Wand (MPI für Informatik – Saarbrücken, DE)

In this talk, I will discuss how to automatically infer rules to build shapes from examples. Given some exemplar geometry, we want to construct a shape grammar that describes a class of objects that are all similar to the input. As a model of similarity, we use local similarity, i.e., local pieces of the output must match the input (as in texture synthesis). The key idea is to examine the symmetry structure of the data in order to find an explicit set of rules that provably constructs only geometry that is similar to the input exemplar.

I will show some result on synthetic 3D meshes as well as scanner data, and conclude the talk with some ideas potential generalizations of the presented framework.

3.34 Implicit scene context for object segmentation and classification

Jan Dirk Wegner (Leibniz Universität Hannover, DE)

Joint work of Wegner, Jan Dirk; Rosenhahn, Bodo; Soergel, Uwe
Main reference Wegner, J.D.; Rosenhahn, B.; Soergel, U., "Implicit scene context for object segmentation and classification," 33rd Annual Symposium of the German Association for Pattern Recognition (DAGM), 2011, accepted for publication.

Our aim is to segment and classify objects in remote sensing images for automatic mapping. A class label is assigned to each pixel. In complex scenes with lots of different object categories like urban areas contextual knowledge may add valuable information if local object descriptors deliver ambiguous results. We learn object-context from the background class of partially labeled images and introduce it as a prior. Local object descriptors and contextual knowledge are combined in a Conditional Random Field framework to label each pixel with the most likely object class. Experiments with simulated data and images of computer vision benchmark data sets, representing context of low and medium complexity, lead to promising results. Context learned from patterns in unlabeled subcategories significantly improves results. Tests with remote sensing data of urban scenes, including context of very high complexity, indicate need for further refinements. More sophisticated contextual learning is necessary to capture complex patterns.

3.35 Inference methods for structure and motion computation: avoiding mistakes before it is too late

Christopher M. Zach (ETH Zürich, CH)

Repetitive and ambiguous visual structures pose a severe problem in many computer vision applications. For instance, erroneously estimated poses between unrelated images with visually similar content can lead to severely distorted 3D models. Our goal is to identify incorrect geometric relations from a set of hypothesized ones, which are typically given as fundamental matrices, homographies, or absolute orientations. Identification of such erroneous relations solely based on low level and local information, e.g. by robust matching techniques and bundle adjustment, is not always possible. The following two cues are helpful to detect incorrect visual relations: (i) determining undetected but predicted visual structures given hypothesized relations, and (ii) verifying the internal consistency of estimated visual relations on a non-local scale. We propose to incorporate these cues particularly into a structure-from-motion framework in order to detect incorrect visual relations, and state this task as Bayesian inference problem. Unlike traditional SfM approaches, where only evidence for the validity of an estimated visual relation is collected, our framework additionally uses indicators explicitly assessing the incorrectness of putative relations. The ultimate goal of this work is obtain a truly incremental, efficient, and fault-tolerant SfM approach.

3.36 Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement

Henning Zimmer (Universität des Saarlandes, DE)

Joint work of Zimmer, Henning; Bruhn, Andres; Luxenburger, Andreas; Weickert, Joachim
Main reference H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement. Computer Graphics Forum," Proc. of Eurographics, vol. 30 (2), pp. 405–414, 2011.
URL http://dx.doi.org/10.1111/j.1467-8659.2011.01870.x

We show how a modern energy-based optic flow method can be used for aligning exposure series used in high dynamic range (HDR) imaging. The main advantage of our approach are the resulting dense displacement fields that can describe arbitrary complex motion patterns, caused by severe camera shake and moving objects.

Additionally, it benefits from several advantages over existing strategies:

(i) It is robust under outliers (noise, occlusions, saturation problems) and allows for sharp discontinuities in the displacement field.

(ii) The alignment step neither requires camera calibration nor knowledge of the exposure times.

(iii) It can be efficiently implemented on CPU and GPU architectures as well as on modern smartphones, e.g. Android phones.

After the alignment is performed, we can additionally use the obtained subpixel accurate displacement fields as input for an energy-based, joint super-resolution and HDR (SR-HDR) approach. It introduces robust data terms and anisotropic smoothness terms in the SR-HDR literature.

4 Working Groups

4.1 Workgroup Summary on Performance Analysis in Dense Correspondence Problems

Daniel Kondermann

The aim of the working group was to discuss how meaningful, objective and accurate information about the performance of dense correspondence methods (DCM) such as optical flow and stereo estimation can be obtained. The group met two times and consisted in total of fourteen members.

Staring out with a brainstorming about applications for DCM, the group quickly found three problem domains.

In the first domain *algorithm characteristics* are of interest. They describe general properties of algorithms such as accuracy, number of parameters, time- and space complexity, graceful degradation, parallelization possibilities, engineerability (ease of full system implementation), confidence estimation and information about alternative solutions in the case of multiple local extrema.

The second problem domain is about *input characteristics*. Any DCM can be applied to any kind of image data. It seems unlikely that an algorithm tuned for particle image velocimetry is capable of dealing with traffic scenes or cinematic movies. On the other hand it is problematic to describe each possible image sequence based on the application it was intended for. Therefore, it would help to find more general descriptors for input data. Two approaches could be of interest: first, global approaches could be used to characterize the similarity of a given scene compared to scenes with known ground truth. This would facilitate the choice of benchmark data to predict the accuracy of a DCM given new data. The second approach could be local: for each finite spatio-temporal neighborhood in a scene, a small dataset such as a structure tensor or any feature descriptor (HoG, FFT-coefficients, SIFT, ...) could be used to characterize the quality of each pixel individually to characterize the scene and possibly predict the quality of the outcome.

The third problem domain being discussed was about publications. Many members of the group agreed that the community currently mainly focuses on accuracy and innovation of DCM. Most other algorithm characteristics and the study of input characteristics are mostly treated with low priority. Based on this "research bias", it seems difficult to advance this field of research. The working group agreed that it would be helpful if the awareness about this bias were increased within the community. This could for example be achieved by conference workshops focusing performance analysis in image processing.

The working group showed great interest in addressing these three problem domains in the future. We hope to establish improved performance analysis methods including more detailed algorithm specifications and new benchmark datasets.

4.2 Workgroup Summary on Challenges in Structure From Motion

Marc Pollefeys

The aim of this working group was to explore open challenges in structure from motion. Structure from motion is the problem of recovering the relative motion/position of the camera(s) as well as the (sparse) 3D structure of the observed scene. While our understanding of this problem has tremendously progressed in the last two decades there are still significant challenges to achieve reliable results on many real world data sets.

First, some argued that the main challenge was to reliably find corresponding points between images. It can indeed often be very difficult to match feature points between images that differ in viewpoint, lighting and camera parameters. This also often depends on the scene which can have limited texture or repeating elements which can lead to a large number of incorrect matches or outliers. The argument was that once potential correspondences had successfully been filtered using multiple view relations, the structure from motion problem itself could be solved.

However, several argued that this was not the case and that state of the art algorithms were often still struggling to achieve good results. Three different types of approach have been proposed. First, sequential structure from motion starts building up the reconstruction by sequentially extending the reconstruction starting from a pair of images and adding images in some order. This approach is often dependent on a good choice for the initial pair of views. To avoid errors accumulating too much, the solution is often globally refined after each new view is added which is very inefficient (and essentially turns the problem from being $O(n^3)$ to $O(n^4)$ with n being the number of images. The second type of approach is hierarchical structure from motion where pairs and triplets of views are first assembled and then further grouped and merged in larger and larger reconstructions. These type of approaches can be more efficient, but also can face challenges when inconsistencies appear between sub-models. The third type of approaches aims to perform batch processing of all the images at once. Factorization approaches are a good example of this. In this case a globally optimal solution can be achieved, but the camera model is strongly simplified and outliers can not be handled. More recent approaches are more general and enforce more robust costfunctions (e.g. L_1), but often face significant computational challenges to scale to large-scale data sets. People also mentioned the recent discrete-continuous approach proposed by Snavely et al. which seemed very promising, but also faced problems of scale and efficiency. This approach alternated between solving a coarse structure from motion problem globally using discrete optimization and refining this solution using non-linear continuous optimization. This was seen as a very promising route, although the current formulation still suffered several important limitations.

All in all, the conclusion of this working group was that although great progress had been made in the last decade or two, no reliable general purpose structure from motion solver that could handle any image collection, even if given potential correspondences containing a reasonable number of correct matches, was yet available.

5 Schedule

Monday, June 25th, 2011

09:15–10:00	Bodo Rosenhahn: *Opening* 2-Minute Self-Presentations
Chair:	Felix Klose
10:15–10:40	Henning Zimmer: *Freehand HDR Imaging of Moving Scenes with Simultaneous Resolution Enhancement*
10:40–11:05	Andres Bruhn: *Structural Prediction for Optical Flow*
11:05–11:30	Daniel Kondermann: *Outdoor Ground Truth for Optical Flow?*
11:30–11:55	Jan Dirk Wegner: *Implicit scene context for object segmentation and classification*
Chair:	Henning Zimmer
14:00–14:25	Stefan Gehrig: *Challenges for Camera-Based Driver Assistance*
14:25–14:50	Friedrich Fraundorfer: *Egomotion estimation and mapping for autonomous systems*
14:50–15:15	Reinhard Klette: *Stereo and motion analysis for vision-augmented vehicles*
Chair:	Falko Schindler
16:05–16:30	Radek Grzeszcuk: *City-Scale Landmark Identification and Text Detection*
16:30–16:55	Torsten Sattler: *Towards Fast Image-Based Localization on a City-Scale*
16:55–17:20	Meinard Mueller: *Realtime 3D Motion Reconstruction from Depth Camera Input*
17:20–17:45	Thomas Helten: *Cross-modal Motion Analysis and Reconstruction*

Tuesday, June 26th, 2011

Chair:	Gerard Pons-Moll
09:15–09:40	Laura Leal-Taixé: *Macro vs micro*
09:40–10:05	Juergen Gall: *Objects are more than bounding boxes*
Chair:	Jan Wegner
10:40–11:05	Ioannis Brilakis: *Achievements and Challenges in Recognizing and Reconstructing Civil Infrastructure*
11:05–11:30	Ralf Dragon: *Towards Feature-Based Situation Assessment for Airport Apron Video Surveillance*
11:30–11:55	Ralf Reulke:*Large Scale Traffic Scene Analysis with Multiple Camera Systems*

Wednesday, June 27th, 2011

Chair:	Jan-Michael Frahm
09:15–09:40	Jan-Michael Frahm: *Efficient Robust Large-scale Reconstruction*
09:40–10:05	Tinne Tuytelaars: *From the lab to the real world: two tales from the road*

Chair:	Ralf Dragon
10:40–11:05	Sameer Agarwal: *Bundle Adjustment in the Large*
11:05–11:30	Frank Dellaert: *Subgraph Preconditioning: The revolutionary new way of using direct graph-based solvers to speed up conjugate gradients*
11:30–11:55	Wolfgang Förstner: *Homogeneity and inhomogeneity of geometric quality in large scale bundle adjustments*
14:00–15:15	Frank & Jan: *Working Group Definition* *Working Group Meetings*

Chair:	Torsten Sattler
15:40–16:05	Silvio Savarese:*Semantic Structure from Motion*
16:05–16:30	Vaclav Hlavac: *Structure in images: Why doesn't it really fly high?*
16:30–16:55	Andreas Schilling: *Measuring and Modeling the World – Bayes and Analysis by Synthesis*
16:55–17:20	Helmut Mayer: *From Image Orientation to Buildings and Trees*

Thursday, June 28th, 2011

Chair:	Thomas Helten
09:15–09:40	Rudolf Mester: *Towards an integrated approach to motion analysis and segmentation*
09:40–10:05	Bärbel Mertsching: *Biologically Inspired Spatiotemporal Saliency Processing for a Computational Model of Visual Attention*

Chair:	Laura Leal-Taixé
10:40–11:05	Michael Goesele: *Workin with Real-World Data*
11:05–11:30	Jean-Sebastian Franco: *Probabilistic methods for shape and motion*
11:30–11:55	Falko Schindler: *Segmentation, Classification and Reconstruction of Surfaces from Point Clouds of Man-made Objects*

Chair:	Dan Raviv
14:00–14:25	Daniel Cremers: *Convex Relaxation Techniques for Geometric Optimization*
14:25–14:50	Christopher M. Zach: *Inference methods for structure and motion computation: avoiding mistakes before it is too late*

Chair:	Christopher Zach
14:50–15:15	Gabriel Brostow: *I see bad pixels, and they don't even know they're bad!*
15:15–15:40	Johan Hedborg: *Rolling shutter video*
16:05–16:30	Dan Raviv: *Affine-invariant diffusion geometry for the analysis of deformable 3D shapes*
16:30–16:55	Michael Wand: *Inverse Procedural Modeling*
16:55–17:20	Gerard Pons-Moll: *Outdoor Human Motion Capture using Data-Driven Manifold Sampling*
17:20–17:45	Thorsten Thormaehlen: *Outdoor Image-based Motion Capture and Reshaping of Humans*

Friday, June 29th, 2011

Chair:	Michael Wand
09:15–09:40	Atsushi Imiya: *Pyramid Transform Revised for Large Sparse and Fast Images and Image Sequences in Real World*
09:40–10:05	Tomas Pajdla: *Robust and Scalable Multi-View Reconstruction*
10:05–10:40	Norbert Krüger: *Two disturbing Remarks on Visual Representations: Hierarchies and Semantics*
10:40–11:05	*Working Group Meeting*
11:05–11:30	*Working Group Meeting*
11:30–11:55	Marc Pollefeys: *Working Group Get Together and Summary of each Group*

Participants

- Steffen Abraham
Robert Bosch GmbH –
Hildesheim, DE
- Sameer Agarwal
Google – Seattle, US
- Ioannis Brilakis
Georgia Inst. of Technology, US
- Gabriel Brostow
University College London, GB
- Andrés Bruhn
Universität des Saarlandes, DE
- Daniel Cremers
TU München, DE
- Frank Dellaert
Georgia Inst. of Technology, US
- Ralf Dragon
Leibniz Univ. Hannover, DE
- Wolfgang Förstner
Universität Bonn, DE
- Jan-Michael Frahm
University of North Carolina –
Chapel Hill, US
- Jean-Sebastien Franco
INRIA Rhône-Alpes, FR
- Friedrich Fraundorfer
ETH Zürich, CH
- Jürgen Gall
ETH Zürich, CH
- Stefan Gehrig
Daimler Res. – Stuttgart, DE
- Michael Goesele
TU Darmstadt, DE
- Radek Grzeszczuk
NRC – Palo Alto, US
- Johan Hedborg
Linköping University, SE
- Christian Heipke
Leibniz Univ. Hannover, DE
- Thomas Helten
MPI für Informatik –
Saarbrücken, DE
- Vaclav Hlavac
Czech Technical University, CZ
- Atsushi Imiya
Chiba University, JP
- Gisela Klette
Auckland University of
Technology, NZ
- Reinhard Klette
University of Auckland, NZ
- Felix Klose
TU Braunschweig, DE
- Reinhard Koch
Universität Kiel, DE
- Daniel Kondermann
Universität Heidelberg, DE
- Norbert Krüger
University of Southern Denmark –
Odense, DK
- Laura Leal-Taixé
Leibniz Univ. Hannover, DE
- Kshitij Marwah
MIT – Cambridge, US
- Helmut Mayer
Universität der Bundeswehr –
München, DE
- Bärbel Mertsching
Universität Paderborn, DE
- Rudolf Mester
Universität Frankfurt, DE
- Meinard Müller
MPI für Informatik –
Saarbrücken, DE
- Thomas Pajdla
Czech Technical University, CZ
- Marc Pollefeys
ETH Zürich, CH
- Gerard Pons-Moll
Leibniz Univ. Hannover, DE
- Dan Raviv
Technion – Haifa, IL
- Ralf Reulke
HU Berlin, DE
- Bodo Rosenhahn
Leibniz Univ. Hannover, DE
- Torsten Sattler
RWTH Aachen, DE
- Silvio Savarese
University of Michigan, US
- Andreas Schilling
Universität Tübingen, DE
- Falko Schindler
Universität Bonn, DE
- Thorsten Thormaehlen
MPI für Informatik –
Saarbrücken, DE
- Tinne Tuytelaars
K.U. Leuven, BE
- Michael Wand
MPI für Informatik –
Saarbrücken, DE
- Jan Dirk Wegner
Leibniz Univ. Hannover, DE
- Christopher M. Zach
ETH Zürich, CH
- Henning Zimmer
Universität des Saarlandes, DE