08291 Abstracts Collection
Statistical and Geometrical Approaches to Visual Motion Analysis
— Dagstuhl Seminar —

Daniel Cremers[1], Bodo Rosenhahn[2] and Alan L. Yuille[3]

[1] Universität Bonn, D
dcremers@cs.uni-bonn.de
[2] MPI für Informatik - Saarbrücken, D
rosenhahn@mpi-inf.mpg.de
[3] Univ. California - Los Angeles, USA
yuille@stat.ucla.edu

Abstract. From 13.07.2008 to 18.07.2008, the Dagstuhl Seminar 08291 "Statistical and Geometrical Approaches to Visual Motion Analysis" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. 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.

Keywords. Motion Segmentation, Statistical Methods, Computational Geometry

08291 – Summary of the Organizers

Daniel Cremers, Bodo Rosenhahn and Alan Yuille

Motion analysis is central to both human and machine vision. It involves the interpretation of image data over time. It is crucial for a range of motion tasks such as obstacle detection, depth estimation, video analysis, scene interpretation, video compression and other applications. Motion analysis is difficult because it requires modeling the complicated relationships between the observed image data and the motion of objects and motion patterns (e.g. falling rain) in the visual scene.

This workshop was focused on critical aspects of motion analysis, including motion segmentation and the modeling of motion patterns. The aim was to gather researchers who are experts in the different motion tasks and in the different techniques used. These techniques include variational approaches, level set

methods, probabilistic models, graph cut approaches, factorization techniques, and neural networks. All these techniques can be subsumed within statistical and geometrical frameworks.

We also involved experts in the study of human and primate vision. Primate visual systems are extremely sophisticated at processing motion so there is much to be learnt from studying them. In particular, we wanted to relate the computational models of primate visual systems to those developed for machine vision. Here, several researchers from the cognitive sciences and biologically inspired vision rounded off the overall high quality of presentations and discussions.

Another important component of the workshop was to develop datasets of image sequences with the associated motion ground truth. These datasets can be used as benchmarks to compare the performance of motion analysis models. They can also be used as data to train statistical models of motion analysis. Datasets with ground truth are being increasingly used in other aspects of machine vision but, at present, there are only very limited motion datasets (e.g. the Yosemite sequence). Here we would like to point out the empeda test sequences, available at `http://www.citr.auckland.ac.nz/6D/` or several discussions about the current optic flow benchmark at middlebury `http://vision.middlebury.edu/flow/`

We made the seminar interactive with plenty of time for discussion. The participants were also encouraged to exchange different modeling techniques and research experiences. We have also identified outstanding unsolved problems in motion analysis and discussed them during highly active group meetings. The participants of this seminar enjoyed the atmosphere and the services at Dagstuhl very much. The quality of this center is unique. It was also a pleasure for the organizers to be involved in a radio interview of the *Computer Club Zwei*. The interview about the Dagstuhl seminar (in German) is available at
`http://www.audioads.de/files/14316/CC-Zwei-118-low.mp3`.

There will be an edited book (within Springer's series on Lecture Notes of Computer Science, LNCS) following the seminar, and all seminar participants have been invited to contribute with chapters. The deadline for those submissions is in November 2008 (allowing to incorporate results or ideas stimulated by the seminar), and submissions will be reviewed (as normal). Expected publication date is 2009.

Tracking closed curves with non-linear stochastic filters

Christophe Avenel (INRIA - Rennes, FR)

The joint analysis of motions and deformations is crucial in a number of computer vision applications. We introduce here a non-linear stochastic filtering technique to track the state of a free curve. The approach we propose is implemented through a particle filter which includes color measurements characterizing the target and the background respectively. The dynamics involved is formulated as a stochastic differential equation. This allows us to get a continuous-time representation of the curve trajectory and, thus, to infer inter-frame deformations.

The curve is defined by an implicit level-set representation and the stochastic dynamics is expressed on the level-set function. It takes the form of a stochastic differential equation with Brownian motion of low dimension. Specific noise models lead to traditional evolution laws based on mean curvature motions, while other forms lead to new evolution laws with different smoothing behaviors. In these evolution models, we propose to combine local motion information extracted from the images and an incertitude modeling of the dynamics. The associated filter we propose for curve tracking thus belongs to the family of conditional particle filters. Its capabilities are demonstrated on various sequences with highly deformable objects.

Keywords: Tracking curves non-linear stochastic filters level-set

Statistical and Numerical Analysis of Image Noise for Optical Flow

John L. Barron (University of Western Ontario, CA)

We present two methods for estimating noise in images (derivative noise, based on high order intensity derivative distributions and model noise, based on the goodness of local least squares fits of a plane to the intensity data) and show how they can be used to reduce optical flow error while maintaining density of the flow. We illustrate our framework on the well-known Lucas and Kanade and Horn and Schunck optical flow algorithms using quantitative average angular error analysis for the Yosemite and Marble synthetic sequences.

Joint work of: Beauchemin, Stephen; Barron, John L.

A Variational Framework for Simultaneous Motion Estimation and Restoration of Motion-Blurred Video

Benjamin Berkels (Universität Bonn, DE)

The problem of motion estimation and restoration of objects in a blurred video sequence is addressed in this paper. Fast movement of the objects, together with the aperture time of the camera, result in a motion-blurred image. The direct velocity estimation from this blurred video is inaccurate. On the other hand, an accurate estimation of the velocity of the moving objects is critical for restoration of motion-blurred video. Therefore, restoration needs accurate motion estimation and vice versa, and a joint process is called for. To address this problem we derive a novel model of the blurring process and propose a Mumford-Shah type of variational framework, acting on consecutive frames, for joint object deblurring and velocity estimation. The proposed procedure distinguishes between the moving object and the background and is accurate also close to the boundary of the moving object. Experimental results both on simulated and real data show the importance of this joint estimation and its superior performance when compared to the independent estimation of motion and restoration.

Joint work of: Bar, Leah; Berkels, Benjamin; Rumpf, Martin; Sapiro, Guillermo

Full Paper:
http://www.ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=SS4409009&isnumber=4408819

Efficient kernel densities for human motion tracking

Thomas Brox (TU Dresden, DE)

Kernel density estimators avoid assumptions on the shape of the approximated distribution. Particularly, they can deal with multimodal distributions. We suggest to model priors on human motion by means of kernel densities, which allows for mixtures of motion patterns. Since human motion describes a low-dimensional manifold in a high-dimensional input space, the performance of the estimator can be improved by using adaptive, anisotropic kernels. We further improve on the standard kernel density estimator by means of a cluster tree structure that allows for efficient access to a set of near neighbors of a query pattern. This reduces the linear time complexity of kernel density estimation to logarithmic complexity. With the estimated densities we can synthesize new motions or support tracking of human motion in video.

Keywords: Human motion, kernel density estimation

Deinterlacing with motion-compensated anisotropic diffusion

Andrés Bruhn (Universität des Saarlandes, DE)

We present a novel deinterlacing scheme that makes consequent use of discontinuity-preserving partial differential equations (PDEs). It combines the accuracy of recent variational motion estimation techniques with the directional interpolation qualities of anisotropic diffusion filters. Our algorithm proceeds in three steps: First, we interpolate the interlaced images by means of a spatial edge enhancing diffusion process (EED). Then we apply the variational optical flow technique of Brox et al. (2004) in order to obtain a precise interframe registration. Finally we use a spatiotemporal generalisation of EED for motion-compensated inpainting of the missing data in the original sequence. Experiments demonstrate that the proposed method outperforms not only classical deinterlacing schemes, but also a recent PDE-based approach.

Dynamical Shape Priors for Implicit Shape Representations

Daniel Cremers (Universität Bonn, DE)

In recent years, researchers have proposed to introduce statistical shape knowledge into level set based segmentation methods in order to cope with insufficient low-level information. While these priors were shown to drastically improve the segmentation of familiar objects, so far the focus has been on statistical shape priors which are *static in time*. Yet, in the context of tracking deformable objects, it is clear that certain silhouettes (such as those of a walking person) may become more or less likely over time. In this talk, I will address the challenge of learning *dynamical* statistical models for implicitly represented shapes. I will show how these can be integrated as dynamical shape priors in a Bayesian framework for level set based image sequence segmentation. We assess the effect of such shape priors "with memory" on the tracking of familiar deformable objects in the presence of noise and occlusion. We show comparisons between dynamical and static shape priors, and between models of pure deformation and joint models of deformation and transformation. Our experiments demonstrate, that level set based segmentation and tracking can be strongly improved by exploiting the temporal correlations among consecutive silhouettes which characterize deforming shapes. In addition, I will show that such problems can be solved in a globally optimal manner using techniques of convex relaxation and Lipschitz optimization.

Keywords: Shape, segmentation, dynamics, convex optimization

Joint work of: Cremers, Daniel; Schmidt, Frank R.; Barthel, Frank

Full Paper:
http://www-cvpr.iai.uni-bonn.de/pub/pub/cremers_et_al_cvpr08.pdf

Full Paper:
http://www-cvpr.iai.uni-bonn.de/pub/pub/cremers_pami06.pdf

Adaptive Dynamic Background Modeling

Gianfranco Doretto (General Electric - Niskayuna, US)

N video surveillance applications, where cameras are typically fixed or panning at constant speed, object tracking heavily exploits the motion cues of the objects in the scene. Such salient motion is normally mixed with nuisance motion, due to scene changes and the possible motion of the camera. Salient motion is typically extracted through a foreground-background segmentation process, which has to be robust to every type of scene changes. State-of-the-art techniques are still unsatisfactory in presence of strong nuisance motion. We describe an approach

that tries to overcome these limitations. We provide a formal definition of what we mean by nuisance, and salient motion, and we formulate the salient motion detection problem as the problem of distinguishing between stationary and non-stationary stochastic processes. This enables building a hierarchy of models that can be learned adaptively online, and that accounts for the stationary nature of the nuisance motion, induced by both the camera and the scene changes. We show how salient motion detection can be cast into a model change detection problem, which we propose to solve optimally online, using the sequential generalized likelihood ratio test. We conclude by showing proof-of-concept results, which demonstrate that the approach has the potential to detect salient motion even in severely cluttered scenes.

2D Human Pose Estimation in TV Shows

Vittorio Ferrari (ETH Zürich, CH)

The goal of this work is fully automatic 2D human pose estimation in unconstrained TV shows and feature films. Direct pose estimation on this uncontrolled material is often too difficult, especially when knowing nothing about the location, scale, and appearance of the person. We propose an approach that progressively reduces the search space for body parts, to greatly facilitate the task for the pose estimator. Moreover, when video is available, we propose methods for exploiting both the continuity of appearance and the continuity of pose for improving the estimation based on individual frames. The latter involves extending the single-frame kinematic graphical model to a full spatio-temporal model over multiple frames, embedding both kinematic and temporal continuity constraints at the same time. The proposed approach is evaluated on four episodes (3 hours) of the TV show Buffy the Vampire slayer. Based on newly produced ground-truth annotation of 2D poses for several hundred frames, we evaluate our approach quantitatively. This ground-truth data has been released on our website www.vision.ee.ethz.ch/ ferrari.

Keywords: Human pose estimation; monocular human tracking; probabilistic inference

Joint work of: Ferrari, Vittorio; Marin-Jiminez, Manuel; Zisserman, Andrew;

Full Paper:
http://www.robots.ox.ac.uk/∼ferrari/Publications/ferrari08cvpr.pdf

Model selection and confidence estimation for image sequence analysis

Christoph Garbe (Universität Heidelberg, DE)

The estimation of complex motion requires refined motion models, linking changes in image brightness to parameters of motion. These models often encompass a large number of degrees of freedom.

The major problem with these complex motion models is that they can be prone to overfitting in the presence of image noise. In order to yield highly accurate estimates of complex motion, the problem of model selection has to be solved. In this contribution we present an approach based on the residuals of the local structure tensor approach to select the optimum model based on the noise level present in the image sequence. The residuals are tested for normality and their deviation are compared to the one expected for a certain noise level. Based on these tests the optimal model can be found. We present results on test sequences, underlining the applicability of this approach.

Apart from model selection, a problem in motion estimation is the identification of areas in which an accurate estimation of motion is not feasible. In this contribution we present two different algorithms than can be applied to motion fields computed from any motion estimation techniques. They rely on learning typical motion fields for a given application. The estimated flow field is then compared to the principal components of typical motion fields and deviances are interpreted as a confidence measure. The first approach is based on heuristics, while a second approach is based on test statistics. Here, the computed p-value is the sought confidence value. These confidence values are compared to previously published techniques on ground truth sequences. Results indicate significant increases of both local and global motion estimation techniques by specifying the flow field based on the proposed confidence measures.

In a post-processing step we perform an interpolation of the specified flow field, leading to dense estimates. These dense flow fields are more accurate than previous estimates, as is shown on ground truth sequences. Thus, in this contribution a framework is presented to increase the accuracy of motion estimation techniques by taking model selection and confidence values into account.

Keywords: Motion estimation, model selection, confidence values

Joint work of: Garbe, Christoph; Kondermann, Claudia; Andres, Björn; Mester, Rudolf

Representation of body motion by learning of primitives in brains and machines

Martin A. Giese (Universitätsklinikum Tübingen, DE)

The learning of complex spatio-temporal patterns provides the basis of the control and recognition of body motion in biological systems. Experimental evidence suggests that biological representations of complex movements are structured in terms of temporal and spatial primitives. The talk gives examples how primitive-based representations of complex body movements can be learned from trajectory data, and how such representations can be exploited for studying visual motion recognition and in technical applications.

Keywords: Body motion, learning, primitives, visual motion recognition, motion synthesis

The representation of space in an expanding room

Andrew Glennerster (University of Reading, GB)

In an immersive virtual environment, observers fail to notice the expansion of a room around them and consequently make gross errors when comparing the size of objects. This result is difficult to explain if the visual system continuously generates a 3D model of the scene based on known baseline information from interocular separation or proprioception as the observer walks. An alternative is that observers maintain a representation like a view graph and have an expectation of the images they will receive but are insensitive to the rate at which images arrive as they walk. I describe the way in which the eye movement strategy of animals simplifies motion processing if their goal is to move towards a desired image and discuss dorsal and ventral stream processing of moving images in that context. Although many questions about view-based approaches to scene representation remain unanswered, the solutions are likely to be highly relevant to understanding biological 3D vision.

Human Measurement of Optic-flow Parameters: Psychophysics and Bayesian Modeling

Norberto Grzywacz (USC - Los Angeles, US)

Yuille and Grzywacz (1998) proposed that humans interpret the retinal optic flow by fitting models of a few elementary components in a Bayesian manner. We tested psychophysically whether this proposal is applicable when combined with the affine-motion models proposed by Koenderink (1986). These models were translation, rotation, expansion/contraction, and deformation. Consistently with these ideas, we found that humans decompose optic flows composed of rotation and expansion/contraction to discriminate finely between angular velocities or rates of expansion. However, humans decompose these motions from translation and deformation incorrectly. Moreover, humans do not use all of the available information to estimate the center of rotation and focus of expansion. And our data show that humans obtain these optic-flow parameters before (in approximately 50 ms) estimating the angular velocity and rate of expansion/contraction (in approximately 150 ms). We developed a successful Bayesian model for these data. Our model postulates that humans would first estimate the center of rotation and the focus of expansion based on "instantaneous" directional fields (instead of full velocity fields). Then, humans would use these estimates to obtain rates of translation, rotation, expansion/contraction, and shear (instead of deformation). Finally, humans would use slow-and-smooth priors to estimate these rates.

Keywords: Optic Flow, Psychophycs, Bayesian Modeling

Joint work of: Grzywacz, Norberto M.; Lee, Kwan J.; Wurfel, Jeffrey D.; Barraza, José F.

Pose primitive based human action recognition in videos or still images

Vaclav Hlavac (Czech Technical University, CZ)

The talk presented a method for recognizing human actions based on pose primitives. In learning mode, the parameters representing poses and activities are estimated from videos. In run mode, the method can be used both for videos or still images. For recognizing pose primitives, we extend a Histogram of Oriented Gradient (HOG) based descriptor to better cope with articulated poses and cluttered background. Action classes are represented by histograms of poses primitives. For sequences, we incorporate the local temporal context by means of n-gram expressions. Action recognition is based on a simple histogram comparison.

Unlike the mainstream video surveillance approaches, the proposed method does not rely on background subtraction or dynamic features and thus allows for action recognition in still images.

Keywords: Motion analysis, activity from video

Joint work of: Thurau, Christian; Hlavac, Vaclav

See also: Thurau C., Hlavac V.: Pose primitive based human action recognition in videos or still images, CVPR 2008, Anchorage, Alaska.

Learning object categories from motion patterns

David C. Hogg (University of Leeds, GB)

Object detection has advanced significantly in the past five years. However, problems arise in treating this visual competence in isolation, particularly in moving beyond a supervised learning regime. The talk will present a partial solution to this problem, through embedding object detection within the wider spatio-temporal context of patterns of activity.

Keywords: Machine learning, computer vision, motion analysis

Joint work of: Hogg, David C.; Murali, Krishna; Cohn, Anthony

Geometrical vs. Statistical Methods for Robot Vision

Atsushi Imiya (Chiba University, JP)

In this talk, I report the recent our results on "a Gibsonian model of the robot navigation."

For the first part, there are strong supports from behavior psychologists. However, for the second part, till now, there is no proof from the view point of psychology. attention to compute the next collision-free direction. Therefore, the robot requires a cyclic mechanism directing attention to the view and computation of the collision-free direction from that view.

We introduce a statistical learning method for the visual navigation of the mobile robot. For the visual navigation of an autonomous robot, the detection of the collision avoidance navigation direction from an image/image sequence captured by imaging systems mounted on the robots a fundamental task. We develop a PCA-based method for the computation of the collision-avoidance navigation-direction from an image sequence captured by a monocular camera system mounted on the robot. The robot detects the free space for the navigation and computes the collision-free direction. For the extraction of the free space, the robot separates the dominant plane and obstacle area using independent component analysis.

For the computation of the collision avoidance direction, our robot computes the principal component of the gradient of the visual potential field which describes the obstacle area. Some experimental results of navigating the mobile robot in synthetic and real environments are presented.

From these results, we arrived the conclusions that

1. The brain does solve mathematics.
2. Mathematics does explain the brain.
3. Mathematics, however, derives the fast algorithms.
4. Computer simulates mathematical models of the brain.

Keywords: Tba

Joint work of: Imiya, Atsushi; Naoya, Ohnish; Yosihiko, Mochizuk

Evaluation of Stereo and Motion Techniques on Real-World Video Sequences

Reinhard Klette (University of Auckland, NZ)

The talk informs about a set of calibrated and geometrically rectified stereo sequences (provided by Daimler A.G. in 2007) and their use for evaluating various techniques for stereo or motion analysis. Ground truth is estimated assuming that the ego-vehicle was driving on a plane. A new technique, combining a KLT tracker with an estimation of 3D directions of motion vectors, is also presented.

Keywords: Performance evaluation, stereo, video, ground truth

Ventricle Segmentation in 3D Image Sequences by using Temporal Coherence and 4D Shape Priors

Timo Kohlberger (Siemens - Princeton, US)

Segmentation of the beating heart in the image sequences, in particular of those of the left ventricle, play an important role in the diagnosis of heart diseases. All major medical limaging techniques like, e.g. Ultrasound, CT, MRI, or Single Photon Emission Computer Tomography (SPECT) are able to acquire 3D images at different time points along a heart cycle. However, state-of-the-art segmentation techniques usually make either no or only local use of the temporal information along the time axis. Instead, in the proposed approach, the information along the temporal dimension is fully taken into account in the segmentation process. That is reached by modeling time as a fourth dimension and to extend each component of the well-known Chan-Vese level set segmentation framework with a statistical shape prior into the four-dimensional spatio-temporal space in a straightforward manner.

By this, a significantly higher robustness of the underlying minimization is reached. Experimental results on Ultrasound and SPECT footage clearly show the significantly improved robustness and reconstruction quality even at the presence of sparse information, high noise and low temporal resolutions.

Keywords: Level set segmentation, spatio-temporal segmentation, ventricle segmentation, PET image processing, SPECT image processing, CT image processing, spatio-temporal priors, statistical shape models

Feature correspondence via graph matching and global optimization

Vladimir Kolmogorov (Univ. College London, GB)

I will consider the problem of establishing correspondences between sparse image features related by an unknown non-rigid mapping and corrupted by clutter and occlusion. A popular approach is to formulate it as a *graph matching* optimization problem. In general, it is NP-hard. I will present a new method which outperforms existing graph matching techniques that we tested, and finds a global minimum in the majority of our examples within a minute.

Joint work with Lorenzo Torresani and Carsten Rother.

Extracting semantic maps using motion information

Norbert Krüger (Univ. of Southern Denmark - Odense, DK)

In this talk, we will present an approach in which maps of the 3D environment become established by making use of motion estimation.

Unlike many other approaches, we do not operate in a 3D point space but on local symbolic entities (called multi-cue primitives). The motion information is used to remove outliers as well as to achieve more precise and complete 3D maps of the environment. These maps allow for semantic reasoning processes which becomes facilitated by the disambiguation process.

The disambiguation process improves the representation in two aspects:

First, it discards erroneous primitives by tracking them over time.

Second, it corrects the primitives' pose over time using independent Kalman filters.

We show results in two rather different scenarions: Firstly, we extract shape representations of objects in a cognitive robot-vision system in the context of the EU project PACOplus. Secondly, we extract 3D maps while driving a car equipped with a stereo system in the context of the EU project Drivsco. As part of this approach, we also give results on motion estimation combining point and line information.

In both scenarios we show, besides the improved reconstructions, the potential of the semantic reasoning on the extracted representations.

Keywords: Motion Estimation, Point and Line Correspondences, Early Cognitive Vision, Semantic Maps

Joint work of: Krüger, Norbert; Pugeault, Nicolas; Pilz, Florian

Recognizing and Synthesizing Human Actions using Parametric Hidden Markov Models

Volker Krüger (Copenhagen Inst. of Technology Aalborg University, DK)

In order to describe non-periodic human arm movements with classical Hidden Markov Models (HMMs), one needs to cope with the different parameterizations the arm actions may have. For example "reaching for an object" is a movement which depends on the location of the object to be grasped. In this talk we introduce parametric HMMs (PHMMs) which allow to represent a particular arm action independent to the possible different parameterizations.

As HMMs are generative models, we use PHMMs for the synthesis and recognition of arm actions. We will show how PHMMs can be use for parameterization invariant action recognition and discuss robustness of the recognition as well as the precision of the synthesis using the PHMMs.

We will present experimental results for human robot interaction where a humanoid robot a) learns the movements from a human, b) executes these movements in completely new scenarios and c) is able to recognize the movements when performed by the human. We conclude with some early experimental results on using PHMMs to improve visual 3D human body tracking.

Keywords: Action representation, action synthesis, action recognition, humanoid robots

Vision-Based Place Recognition and Cognitive Mapping

Hanspeter Mallot (Univ. Tübingen, DE)

Spatial longterm memory in animals and humans can be modelled as a graph structure with nodes representing memorized places, and edges representing knowledge about navigational transitions between such places. In the talk, two problems of this approach are addressed, (i) how do humans recognize places?, and (ii) how can metric information be added to the graph scheme? In behavioural experiments with human subjects navigating a virtual environment, we showed that featureless image information, as provided by a smooth circular colour gradient in a circular room, suffices to build recognizable place representations. The navigational errors depend on colour contrast and local image variation, as is predicted from plain image difference (or "snapshot") models as have been suggested for insect navigation.

Links of the space graph represent navigational actions or motor programs and will therefore often contain implicit metric information on place-to-place distances and turns. We present an extended multidimensional scaling algorithm for metric embedding of space graphs, i.e. for calculating global place coordinates from place distances and angles in chains of three places. The relevance of this scheme for human spatial memory is discussed.

In conclusion, human spatial memory can be nicely modelled by the topological approach, while global metric knowledge is often poor.

Keywords: Spatial memory, place recognition, space graph, metric maps

Fast and Accurate Optical Flow Estimation: Primal-Dual Schemes and Second Order Priors

Thomas Pock (Universität Bonn, DE)

This talk is concerned with the fast and accurate estimation of optical flow. In the first part of the talk we propose a fast numerical algorithm to compute the minimizer of an energy functional composed of Total Variation regularization and a L1 data term. The resulting numerical scheme can be effectively accelerated by a GPU-based implementation leading to a real-time performance of 30 frames per second for 640x480 images. In the second part of the talk we propose a novel prior term based on second order derivatives. In particular we propose a novel second order operator which allows to separate the magnitude of the second order derivatives from local orientation and shape. We demonstrate the superior performance of the novel second order prior on the Middlebury benchmark data sets for optical flow.

Keywords: Optical Flow, Variational Methods, Realtime, GPU

Joint work of: Pock, Thomas; Cremers, Daniel

Modeling and Tracking Constricted Kinematic Chains

Bodo Rosenhahn (MPI für Informatik - Saarbrücken, DE)

The presentation deals with modeling, markerless tracking and animation of constricted kinematic chains, e.g. athletes interacting with sports gear or people interacting with the environment.

In contrast to classical markerless tracking, the modeling of external constraints during motion capture allows to reduce the search space to a desired manifold which again helps to avoid local minima and resolve ambiguities during tracking. The improved tracking results are also reflected in more realistic animations.

Experimental results on several scenarios show the general applicability of our approach.

Discrete-Continuous Optimization for Optical Flow Estimation

Stefan Roth (TU Darmstadt, DE)

Accurate estimation of optical flow is a challenging task, which often requires addressing difficult energy optimization problems. To solve them, most top-performing methods rely on continuous optimization algorithms. The modeling accuracy of the energy in this case is often traded for its tractability. This is in contrast to the related problem of narrow-baseline stereo matching, where the top-performing methods employ powerful discrete optimization algorithms such as graph cuts and message-passing to optimize highly non-convex energies.

In this talk I will demonstrate how similar non-convex energies can be formulated and optimized discretely in the context of optical flow estimation. Starting with a set of candidate solutions that are produced by fast continuous flow estimation algorithms, the method iteratively fuses these candidate solutions by the computation of minimum cuts on graphs. The obtained continuous-valued fusion result is then further improved using local gradient descent. Experimentally, we demonstrate that the energy formulation is an accurate model and that the proposed discrete-continuous optimization scheme not only finds lower energy solutions than traditional discrete or continuous optimization techniques, but also leads to flow estimates that outperform the current state-of-the-art.

Joint work of: Lempitsky, Victor; Roth, Stefan; Rother, Carsten

A Rudin Osher Fatemi Approach to Optical Flow

Martin Rumpf (Universität Bonn, DE)

The estimation of motion in an image sequence is a fundamental task in image processing.

Frequently, the image sequence is corrupted by noise and one simultaneously asks for a restored sequence and the underlying motion field. In smoothly shaded regions of the restored image sequence the brightness constancy assumption along motion paths leads to a pointwise differential condition on the motion field.

At object boundaries which are edge discontinuities both for the image intensity and the motion field this condition is no longer well-defined. In this paper a total variation type functional is discussed for the joint image restoration and motion estimation.

This functionals appears to be not lower semicontinuous. By the general theory on vector valued total variation functionals its relaxation comes along with a singular part of the energy density which is given by the solution of a local minimization problem at edges. In the case of one space dimension this energy density can be identified explicitly. This relaxation leads to a generalized variational formulation of the brightness constancy assumption.Furthermore, it enables to resolve an ambiguity problem in the selection of foreground and background motion In case of fast motion a minimization of the functional surprisingly favors microstructures at object boundaries, which can be avoided by a suitable scaling of the model.

Joint work of: Rumpf, Martin; Corti, Sergio

An affine optical flow model for dynamic surface reconstruction

Hanno Scharr (Forschungszentrum Jülich, DE)

We presented a differential, affine optical flow like model for multi-camera image sequences. Together with a total least squares estimation scheme it allows to calculate 3d motion and shape of inclined local surface patches. The model is derived by projecting a surface patch model onto to the image planes of a camera grid using a pinhole camera model. Differentials of the thus derived image coordinates of surface patches are then substituted into a differential formulation to the well known brightness constancy assumption for optical flow. In our case, brightness is not only assumed to be constant with respect to time, but also from camera to camera as it is common in structure from camera motion approaches. Therefore parameter estimation differs from typical estimation schemes for optical flow by suitable disentangling of surface and motion parameters from estimated mixed model parameters. Experiments using synthetic image sequences with ground truth available for all estimated surface patch parameters show the validity of the derived model. Results on real multi-grid camera data demonstrate its applicability.

Keywords: Camera grid, multi-camera stereo, 3D motion, surface normals, brightness changes

Joint work of: Schuchert, Tobias; Scharr, Hanno

Handling Occlusions in Region Based Pose Tracking by Increasing the Number of Regions

Christian Schmaltz (Universität des Saarlandes, DE)

Despite great progress achieved in 3-D pose tracking during the past years, handling self-occlusions is still an open issue. This is particularly true in region-based pose tracking, in which problems can even occur when one part of an object is completely in front of another. Using multiple cameras or motion priors was proposed to overcome this problem. However, using multiple cameras is not always sufficient, and when using motion priors, the results strongly depend on choosing the right priors, which might not be readily available. Here, we will model self-occlusions by minimising a joint energy function over multiple objects with multiple components for which more than one silhouette is required during minimisation. This makes it possible to deal with self-occlusions between different parts of the same object as well as with occlusions that occur between different objects. The results we present for simulations and real-world scenes demonstrate the improvements achieved in monocular and multi-camera settings.

Keywords: Pose estimation, tracking, kinematic chain, computer vision, human motion analysis, self-occlusion

Joint work of: Schmaltz, Christian; Rosenhahn, Bodo; Brox, Thomas; Weickert, Joachim; Wietzke, Lennart; Sommer, Gerald

On Image Flows and TV Regularization

Christoph Schnörr (Universität Heidelberg, DE)

The talk will focus on the Rudin-Osher-Fatemi model of image denoising and decomposition and two variations of it. The first one takes into account dominant orientations of the 3D scene structure in order to avoid the well-known "staircasing effects" for smooth vector fields with non-vanishing gradients. The second one provides an approximation of the non-binary image labeling problem that, for instance, is applicable to motion features, with the uniform metric between labels as prior.

Preliminary numerical experiments indicate the performance of the respective global optima for noisy input data.

Keywords: Image denoising, image decomposition, TV regularization, convex programming

High-resolution Motion Layer Decomposition

Thomas Schoenemann (Universität Bonn, DE)

In this talk we present and discuss two alternatives to the traditional way of motion estimation. We focus on inferring scene structure rather than tracing points over time.

We present a framework for real-time motion segmentation which allows to indentify the objects in the scene. This is done based on the motion of the objects and relies on graph cut segmentation.

The second approach, layer decomposition, allows to naturally model and handle occlusion and infer a scene model. Here we contribute both on atheoretical and an applied level. We show how to obtain fine-detailed, high-resolution layer images based on super resolution. We also present regularity terms constraining the intensities and geometry of each layer. Moreover, we present an algorithmic framework to (locally) minimize the arising energy.

This is joint work with Daniel Cremers.

Keywords: Motion analysis, layer decomposition, segmentation

Efficient Dense Scene Flow from Sparse or Dense Stereo Data

Tobi Vaudrey (University of Auckland, NZ)

I presented an approach for estimating the three-dimensional velocity vector field that describes the motion of each visible scene point from a stereo sequence. This vector field is also referred to as scene flow. The technique presented uses only two consecutive image pairs from a stereo sequence. I will cover the differences between 2D optical flow and 3D scene flow. Our main contribution is to decouple the position and velocity estimation steps, and to estimate dense velocities using a variational approach in near real-time. The two-dimensional optical fl ow field is extended into the third dimension by additionally estimating the disparity change between consecutive image pairs. Furthermore, we enforce the scene fl ow to yield consistent displacement vectors in the left and right image pairs. The decoupling strategy has two main advantages: Firstly, we are independent in choosing a disparity estimation technique, which can yield either sparse or dense correspondences, and secondly, we can achieve frame rates of 5 fps on standard consumer hardware (previously this was limited to hours of computational time). The approach provides dense velocity estimates with accurate results at distances up to 50 meters. The velocity field remains dense, even if sparse disparity or occlusions are provided, due to the variational approach.

Keywords: Stereo, optical flow, scene flow, real-time algorithm

Joint work of: Vaudrey, Tobi; Wedel, Andreas; Rabe, Clemens; Brox, Thomas; Cremers, Daniel; Franke, Uwe

Segmentation of Moving Obstacles via Graph Cut

Andreas Wedel (Daimler AG, DE)

Autonomous collision avoidance in vehicles requires an accurate seperation of obstacles from the background, particularly near the focus of expansion. We tackle this problem using fast segmentation of stationary obstacles from video recorded by a single camera that is installed in a moving vehicle. The input image is divided into three motion segments consisting of the ground plane, the background, and the obstacle. This constrained scenario allows for good initial estimates of the motion models, which are iteratively refined during segmentation. The remaining binary partitioning problem is solved by a graph cut on the motion-compensated difference images. Our experimental evaluation shows that the proposed approach leads to fast and accurate obstacle segmentation and distance estimation without prior knowledge about the size, shape or base point of obstacles. We also present a binocular version of the approach which is able to segment moving obstacles in distances up to 70m.

Keywords: Object Segmentation, Graph Cut, Intelligent Vehicles

The Conformal Monogenic Signal of Image Sequences

Lennart Wietzke (Christian-Albrecht-Universität zu Kiel, DE)

This work presents a novel rotational invariant quadrature filter approach - called the 3D conformal monogenic signal - for analyzing local features of any curved 3D signal such as hyperplanes and spheres without the use of steering. The conformal monogenic signal contains the monogenic signal as a special case and combines monogenic scale space, phase, direction/ orientation, energy and curvature in one unified algebraic framework.

The conformal monogenic signal will be theoretically illustrated and motivated in detail by the relation of the 4D Radon transform and the generalized Hilbert transform on the unit sphere. The main idea is to lift up 3D signals to the 4D conformal space where the signal features can be analyzed with more degrees of freedom. Results of this work are the low computational time complexity, the easy implementation into existing Computer Vision applications and the numerical robustness of determining curvature without the need of any derivatives.

Keywords: Monogenic Signal, Local Image Sequence Analysis, Conformal Space, Signal Processing

Joint work of: Wietzke, Lennart; Sommer, Gerald; Schmaltz, Christian

Model selection and velocity estimation using novel priors for object motion

Alan L. Yuille (Univ. California - Los Angeles, US)

Psychophysical experiments show that humans are better at perceiving rotation and expansion than translation. These findings are inconsistent with standard models of motion integration which predict best performance for translation. To explain this discrepency, our theory formulates motion perception at two levels of inference: we first perform model selection between competing models (e.g. translation, rotation, and expansion) and then estimate the velocity using the selected model. We define novel prior models for smooth rotation and expansion using techniques similar to the slow-and-smooth model (e.g. Green functions of differential operators). The theory gives good agreement with the trends observed in human experiments.

Keywords: Velocity estimation, model selection, Green functions, Psychophysics

Joint work of: Yuille, Alan L.; Shaung Wu and HongJing Lu