Computer Vision in Camera Networks for Analyzing Complex Dynamic Natural Scenes

Joachim Denzler

Chair for Computer Vision, Friedrich Schiller University of Jena
07743, Jena, Ernst-Abbe-Platz 2, Germany
denzler@informatik.uni-jena.de

Sensor or camera networks will play an important role in future applications, from surveillance tasks for workplace safety or security in general, over driver assisting systems in automotive and last but not least intelligent homes or assisted living for the elderly. Computer vision in sensor or camera networks defines a couple of currently unsolved problems. First of all, how can we calibrate cameras distributed arbitrarily in the scene without placing artificial or natural calibration patterns in the scene? Second, how do we select and fuse the information provided by different, also multimodal sensors to solve a given problem? Finally, can we handle reconstruction, recognition and tracking tasks in complex and highly dynamic natural scenes which are in almost all cases the environment camera networks are designed for?

The chair for Computer Vision of the Friedrich Schiller University of Jena is working since more than four years on different topics strongly related to future applications in sensor or camera networks. The basis of all investigations is a distributed computer vision system, consisting of six computer controlled pan/tilt/zoom cameras each connected to a high-performance PC, a mobile platform with stereo and PMD range camera, and a high performance image acquisition and processing server. The setup allows for local image processing as well as fusion at a central place. In addition, a computer steerable robot arm is available for recording of images under controlled conditions, benchmarking and ground truth data creation.

In this seminar we report about different aspects of and solutions to computer vision problems with direct applications in camera networks or processing of data of complex and/or natural scenes. In the work of Kähler an online structure from motion (SFM) approach is developed which differs from previously published work in the way feature tracking and 3d reconstruction is handled. Instead of tackling SFM by two steps (tracking and reconstruction), a combined optimization approach is used, where under weak assumptions about the scene the camera parameters as well as the structure of the scene is estimated within one optimization step. Efficient implementation allows for close to real time frame rate processing of input data. Self-calibration of a camera network is the topic of the work of Bajramovic. He presents an optimization approach based on a quality criterion of relative pose estimates as uncertainty measure. Optimization is reduced to shortest path search in a so called camera dependency graph constructed from local relative pose estimates. Closely related to the work of Bajramovic is the question how to identify cameras that share a common view

Dagstuhl Event Proceedings 08422
Computer Vision in Camera Networks for Analyzing Complex Dynamic Natural Scenes
http://drops.dagstuhl.de/opus/volltexte/2009/1865

of the scene. Brückner presents an approach where based on epipolar constraints and RANSAC, common parts are identified in a probabilistic framework.

The work of Munkelt and Trummer introduces solutions to the next best view planning for 3d reconstruction. Instead of taking arbitrary views of an object the best views shall be taken with respect to a predefined quality criterion. While Munkelt reports about a methodology for evaluating next best view algorithms in the context of active illumination, Trummer improves feature point tracking in such applications by incorporating knowledge about the camera movement. As a result an extension of the well known KLT tracker is available which shows better performance not just for the tracking in the image plane itself but also for the 3d reconstruction of an object based on the tracked features.

More related to machine learning and object recognition but nevertheless also important for processing and analyzing complex natural scenes are the results of Hegazy and Rodner. Hegazy has developed a system for generic object recognition based on different local features and boosting for classification. Her results show that combination of optimal features and boosting techniques provides best recognition results. She also demonstrates that her approach is general in the sense that also multimodal data can be combined, currently features from color cameras and range information from PMD devices. In total, the recognition rates can be further improved using such multimodal data analysis. Rodner tackles the question of how to learn object classes from few examples, a problem which occurs in many industrial tasks where training data acquisition is very expensive or recording of the necessary training set is even impossible at all. The main idea of his approach is to combine principles from knowledge transfer and decision trees to use information from so called support classes to regularize randomized decision trees in the case of few training data. Finally, the work of Platzer demonstrates how model based approaches can solve computer vision problems in the case of difficult recording conditions or image quality. In her work the question is how to identify defects in wire ropes automatically from gray value images taken by a line camera. Since defects are even difficult to detect by humans and only few positive samples are available due to the effect of the strong regulations of the respective European norm, standard machine learning techniques fail to solve this problem. She presents a model based approach with the key idea to detect abnormalities from the predicted view of the rope by an analysis-by-synthesis approach.

More related to understanding the function of the human visual system is the work of Koch. He investigates properties of images of different categories with respect to features which are unique for natural scenes. The goal is to identify common features in art and natural scenes which relate the processing of such data within the human visual systems to esthetic perception. The expectation is that understanding the processing principles being the basis of our visual system allows for the design of computer vision algorithms in areas, where currently the machine does not reach the performance of humans. Such areas might be generic object recognition, use of context in computer vision or learning from few examples.