

Workshop Series: Perspectives on Vision Networks (Ed. 2)

Wireless Sensor Networks Lab, Stanford University

Wed. Nov. 28, 2007, 12:00 – 16:30

Packard 204

Organizer: Hamid Aghajan (hamid AT wsnl.stanford.edu)

Agenda in Overview:

12:00 - 12:40 Welcome and introductions

12:40 - 14:00 Talks (20 minutes each)

14:00 - 14:20 Break

14:20 - 16:00 Talks (20 minutes each)

16:00 - 16:30 Open discussion and brainstorming

Agenda in Detail:

12:00 – 12:40 Welcome and introductions

12:40 – 13:00 Wannes van der Mark (TNO, The Hague, The Netherlands), “3-D Scene Reconstruction with a Handheld Stereo Camera”

13:00 – 13:20 Marleen Morbee (University of Ghent, Belgium), “A Distributed Coding-based Content-Aware Multi-View Video System”

13:20 – 13:40 Huang Lee (Stanford University, USA), “Energy-Efficient Occupancy Reasoning in Multi-Cluster Camera Networks”

13:40 – 14:00 Tao Xu (Stanford University, USA), “Coupled Markov Model Based Human Behavior Interpretation in Distributed Vision Networks”

14:00 – 14:20 Break

14:20 – 14:40 Chen Wu (Stanford University, USA), “Vision Algorithm Design Strategies in a Camera Network and its Application in Human Pose Estimation”

14:40 – 15:00 Linda Tessens (University of Ghent, Belgium), Camera Selection in Distributed Video Coding Networks

15:00 – 15:20 Itai Katz (Stanford University and NIST, USA), “Context-based Scene Interpretation in Vision Networks”

15:20 – 15:40 Tommi Maatta (Tampere University, Finland, and Philips Research, The Netherlands), “Multiview Volumetric Reconstruction with Shape-from-Silhouette Method and its Application”

15:40 – 16:00 Nan Hu (Stanford University, USA), “Intelligent Traffic Interpretation with Camera Networks”

16:00 – 16:30 Open discussion and brainstorming

Abstracts:

- **Wannes van der Mark (TNO, The Hague, The Netherlands), “3-D Scene Reconstruction with a Handheld Stereo Camera”**

There are many applications that require accurate three-dimensional (3-D) computer models of real world scenes. Example applications can be found in crime scene investigation, engineering, construction work, and the entertainment industry. In this talk we present a method for obtaining 3-D models of real world scenes. The idea is that a handheld stereo camera is used to capture images of a scene from different viewpoints. A scene model can then be build automatically by merging the resulting 3-D stereo measurements based on the estimated camera ego-motion. In order to reduce complexity, a novel selection process is used to remove very similar images at the beginning of the reconstruction process. The stereo camera ego-motion is then recovered with estimation methods that are robust to various noise influences and outliers. Because the ego-motion estimates are relative, the bundle adjustment technique is applied to reduce error accumulation in the estimated camera trajectory. We will conclude by explaining the methods used for 3-D surface approximation and present an example 3-D reconstruction of a (simulated) crime scene.

- **Marleen Morbee (University of Ghent, Belgium), “A Distributed Coding-based Content-Aware Multi-View Video System”**

Flexible camera networks that do not need wires for communication and power supply, can be of great advantage in many applications, such as discrete crime scene observation, remote campsite surveillance, etc. However, the limited power resources and the use of wireless communication require the development of well-considered algorithms for the data processing and transmission in these networks. In this presentation, we propose an efficient and content-aware data reduction and compression method for a flexible wireless camera network. The framework consists of a central processor and camera, completed by a number of smart Wyner-Ziv cameras. The latter ones provide a content-aware representation of their viewpoint, thus greatly reducing the amount of data to be sent to the central processor. The remaining image data is compressed by employing Distributed Video (DV) coding, i.e. joint decoding of the independently encoded frames of the different cameras, which allows us to achieve good coding efficiency without inter-camera communication.

- **Huang Lee (Stanford University, USA), “Energy-Efficient Occupancy Reasoning in Multi-Cluster Camera Networks”**

Occupancy reasoning is one central task in civilian surveillance systems. During the reasoning, data from the wireless nodes in the network is periodically collected and sent back to base stations for further processing. Because the collection task is performed very frequently, its energy consumption forms a major part of the overall energy consumption in the network. Therefore, designing an energy-efficient data collection algorithm is essential for achieving a long battery lifetime for the network. In this talk, we discuss the design of such data collection process in stationary multi-cluster wireless camera networks and present a practical and energy-efficient solution. We assume that the nodes in the network are divided into disjoint clusters based on the geographic locations. In each cluster, the node data is first aggregated and then the cluster-head is responsible for sending the compiled data back to the base stations. The energy-efficiency is achieved by combining the following design methods: (a) Adopting network optimization techniques, the optimal data collection scheme, including the cluster-head selection, considering the balance of the sensing and communication energy among all the nodes in the network is calculated. This gives the lower bound for the energy required for the data collection process. (b) Instead of using a fixed network topology (communication tree), a set of trees are constructed and the communication tree varies over different data collection cycles. We show that this method achieves an

average energy consumption rate very close to the optimal value. (c) A data aggregation method is proposed such that all the nodes transmit equal amount of data and the size of transmission packets is minimized to reduce the energy consumption. The performance of the proposed solutions is evaluated through simulations.

- **Tao Xu (Stanford University, USA), “Coupled Markov Model Based Human Behavior Interpretation in Distributed Vision Networks”**

Human behavior interpretation is an essential step in assisted living and monitored care applications. In this talk, we propose a Coupled Markov Model (CMM) based framework to interpret the observed human behavior real-time. The collaboration and information fusion between cameras in a distributed vision network are first quantitatively described by the CMMs which are then trained to represent five different human behaviors considered in this work. Finally, human behavior interpretation is performed using a Bayesian framework based on Maximum Likelihood (ML) method. The experimental results demonstrate the superior performance of our proposed approach.

- **Chen Wu (Stanford University, USA), “Vision Algorithm Design Strategies in a Camera Network and its Application in Human Pose Estimation”**

There are two distinguishing features in the design of vision algorithms in a camera network. The first one is to reduce image data in local processing, so that the output of a single camera can be shared in the network. The second is to fuse observations from different views to perform what cannot be done from monocular cameras and to achieve robustness. Combining with other approaches for vision problems, we believe that optimal fusion of three dimensions (image features, space and time) would be the solution for a versatile and robust vision system of a camera sensor network. A method for human pose estimation is presented exploiting the above concepts. The first novelty roots in the basic requirement from the network, i.e., to reduce image data into short feature descriptions, while keeping enough information for the collective 3D reconstruction between cameras. The second novelty is the attempt to use different image features for different body parts. The human body presents complex and changing appearance in the images, thus it is very difficult to use a single image feature to describe all the variety. We choose to use different image features to represent different body parts more effectively. 3D pose estimation results for a boxing sequence will be shown.

- **Linda Tessens (University of Ghent, Belgium), “Camera Selection in Distributed Video Coding Networks”**

Within a camera network, the contribution of a camera to the observation of a scene depends on its viewpoint and on the scene configuration. This is a dynamic property, as the scene content is subject to change over time and the camera configuration might not be fixed, e.g. in a mobile network. An automatic selection of a subset of cameras that significantly contributes to the desired observation of a scene can be of great value for the reduction of the amount of transmitted or stored image data. We apply this concept to a wireless camera network that encodes the captured image data using distributed video coding (DVC) principles. In such an environment, particular attention needs to be paid to the distribution of the different camera roles within the network.

- **Itai Katz (Stanford University and NIST, USA), “Context-based Scene Interpretation in Vision Networks”**

In this talk we present a framework for improving object classification performance using contextual information. Our method uses probability maps to guide classifiers to image regions likely to contain the

object in question, based on the object's past positions and the positions of surrounding objects. We contrast our method with a baseline unguided classifier and show that using probability maps as a pre-processing step significantly reduces the number of positions a classifier needs to evaluate.

- **Tommi Maatta (Tampere University, Finland, and Philips Research, The Netherlands), “Multiview Volumetric Reconstruction with Shape-from-Silhouette Method and its Application”**

System goal is a home-to-home visual communication and awareness by combining video processing and lighting to generate shadow-like characters. A multi-camera system is used to capture a scene from different viewpoints. Through silhouette extraction, skin part detection and SFS a volumetric reconstruction of the person is built. Resulting visual hull is used for visualization and as input for body model fitting process. With this 3D approach more interactive experience can be achieved.

- **Nan Hu (Stanford University, USA), “Intelligent Traffic Interpretation with Camera Networks”**

In this project, we are interested in providing intelligence for a particular smart environment – roadway traffic. The proposed framework for traffic interpretation is made up of a dynamic vision sensor network with each node being a smart camera, i.e. a physical camera with an individual software agent. Thus, each smart camera has the ability of processing vision and reasoning task locally. Through the communication within the network, evidences are shared among camera-agents to enhance the reasoning, hence achieving a better understanding of the traffic model and the causes of flow variation. Within the framework, both the on-camera vision and reasoning and the inter-camera communication schemes are proposed and described in detail.