Workshop on

Multi-camera and Multi-modal Sensor Fusion: Algorithms and Applications (M^2SFA^2 2008)

Marseille, France (October 18, 2008) - in conjunction with ECCV 2008

ABSTRACTS

Program at a glance	
08:30 - 08:50	Posters set-up
08:50 - 09:00	**Welcome and Introduction**
09:00 - 09:40	**Oral session**: *Multi-modality: video + depth*
09:40 - 10:30	***Keynote presentation***: *Nils Krahnstoever, GE Global Research*
10:30 - 11:00	Coffee
11:00 - 12:20	**Oral session**: *Algorithms: multi-sensor calibration and control*
12:20 - 14:00	Lunch
14:00 - 15:30	**Poster session**
15:30 - 16:00	Coffee
16:00 - 17:00	**Oral session**: *Multi-camera networks: applications*
17:00 - 17:40	***Panel***: *Application-driven design of multi-camera systems*
17:40 - 18:30	***Group discussion***: *Opportunities in multi-sensor research: driven by concept or technology?*

General chairs: Andrea Cavallaro (*Queen Mary, U. of London, UK*), Hamid Aghajan (*Stanford University, USA)*

Program committee: Francois Bremond (*INRIA, France*); Josep Casas (*UPC, Spain*); Tanzeem Choudhury (*Dartmouth College, USA*); Maurice Chu *(PARC, USA)*; C. De Vleeschouwer *(UCL, Belgium)*; Pascal Frossard *(EPFL, Switzerland)*; Luis Matey (*CEIT, Spain*); Jean-Marc Odobez (*IDIAP, Switzerland*); James Orwell (*Kingston U., UK*); Wilfried Philips (*U. of Gent, Belgium*); Ronald Poppe (*U. of Twente, Netherlands*); Fatih Porikli (*MERL, USA*); Carlo Regazzoni (*U. of Genoa, Italy*); Rainer Stiefelhagen (*U. of Karlsruhe, Germany*); Ming-Hsuan Yang (*Honda Research, USA*); Li-Qun Xu (*BT, UK*).

Workshop webpage: http://www.elec.qmul.ac.uk/staffinfo/andrea/M2SFA2.html

Sponsors:

MITSUBISHI ELECTRIC RESEARCH LABORATORIES

SEVENTH FRAMEWORK PROGRAMME

apidis

Oral sessions

Multi-modality: video + depth

09:00 A Unified Approach to Calibrate a Network of Camcorders and ToF cameras

Li Guan (University of North Carolina-Chapel Hill, USA); Marc Pollefeys (ETH Zurich, Switzerland)

In this paper, we propose a unified calibration technique for a heterogeneous sensor network of video camcorders and Time-of-Flight (ToF) cameras. By moving a spherical calibration target around the commonly observed scene, we can robustly and conveniently extract the sphere centers in the observed images and recover the geometric extrinsics for both types of sensors. The approach is then evaluated with a real dataset of two HD camcorders and two ToF cameras, and 3D shapes are reconstructed from this calibrated system. The main contributions are: (1) We reveal the fact that the frontmost sphere surface point to the ToF camera center is always highlighted, and use this idea to extract sphere centers in the ToF camera images; (2) We propose a unified calibration scheme in spite of the heterogeneity of the sensors. After the calibration, this multi-modal sensor network thus becomes powerful to generate high-quality 3D shapes efficiently.

09:20 Integrating Visual and Range Data for Robotic Object Detection

Stephen Gould (Stanford University, USA); Paul Baumstarck (Stanford University, USA); Morgan Quigley (Stanford University, USA); Andrew Ng (Stanford University, USA); Daphne Koller (Stanford University, USA)

The problem of object detection and recognition is a notoriously difficult one, and one that has been the focus of much work in the computer vision and robotics communities. Most work has concentrated on systems that operate purely on visual inputs, i.e., images, and largely ignores other sensor modalities. However, despite the great progress made down this track, the goal of high accuracy object detection for robotic platforms in cluttered real-world environments remains elusive. Instead of relying on information from the image alone, we present a method which exploits the multiple sensor modalities available to a robotic platform. In particular, our method augments a 2-d object detector with 3-d information from a depth sensor to produce a "multi-modal object detector." We demonstrate our method on a working robotic system and evaluate its performance on a number of common household/office objects.

09:40 Keynote presentation: Nils Krahnstoever, GE Research
Applications and challenges for multi-camera and multi-sensor vision systems

This talk will give an industry perspective on multi-camera and multi-sensor computer vision systems and challenges. We will discuss several example applications and draw on experience gained from several of our industrial research efforts in retail, mass-transit, perimeter protection and broadcast. In particular, we will discuss GE Global Research's intelligent video framework targeted at the automatic surveillance of large sites and discuss challenges around calibration, detection and tracking, PTZ control, and biometrics. We will also discuss the need for robustness, scalability, ease of deployment and cost issues in today's security systems, which are topics of particular importance as vision technologies are transitioning from high-end to main-stream security markets. As part of the talk we will discuss the challenges faced by vision system scientists and developers in the industry and suggest future research directions for the computer vision community.

Bio: Dr. Krahnstoever is a senior research scientist at the Visualization and Computer Vision Lab at GE Global Research. At GE he has led and performed research for GE Security, NBC Universal, Lockheed Martin, and many other GE businesses and customers. His current work is centered around large-scale real-time surveillance systems with a particular focus on person detection and tracking, behavior recognition, calibration, search, embedded systems and performance evaluation. Other research interests include vision-based special effects in broadcast and human computer interaction. Dr. Krahnstoever is currently a PI on a research effort with the Sensor and Surveillance unit of the National Institute of Justice. While at GE, Dr. Krahnstoever has led a number of successful product and system deployments such as vision-based special effects for Sunday Night Football on NBC and several successful security system deployments.

Dr. Krahnstoever has more than ten years of experience in applied computer vision research. Prior to joining GE in 2003 he obtained a M.S. in Physics from the University of Kiel, Germany in 1998. Following graduation he worked for PHILIPS Medical Systems as an intern. He obtained his Ph.D. in Computer Science from the Pennsylvania State University in 2003. During his Ph.D. he worked as the Director of Computer Vision for Advanced Interfaces Inc. (now VideoMining).

Dr. Krahnstoever regularly serves as a referee for journals and conferences and published over 35 peer reviewed scientific publications and patents.

Algorithms: multi-sensor calibration and control

11:00 Automatic Calibration of Camera Networks based on Local Motion Features

Keni Bernardin (Universität Karlsruhe, Germany); Cem Aslan (Universität Karlsruhe, Germany); Rainer Stiefelhagen (Universität Karlsruhe, Germany)

This paper introduces a new technique to automatically calibrate the extrinsic parameters of a network of cameras without using dedicated calibration objects or markers. Instead, the motion of persons walking naturally through a scene is used. Simple foreground and motion features are extracted from the individual image sequences. A Hough transform is applied in a specially defined parameter space to estimate the relative geometry between camera pairs without solving the problem of finding correct feature correspondences between views. All possible feature correspondences are examined, and a modified gradient descent algorithm is used to find the set of optimum calibration parameters within the Hough space. After calibrating each camera pair the resulting camera network is built up using a global error minimization technique. The approach is tested on several indoor scenarios and shows a high degree of robustness, especially when multiple persons enter the scene, making it difficult to resolve feature correspondences. The correctness and precision of individual camera calibrations and of the resulting camera network are thoroughly evaluated, showing that triangulation errors as low as 5cm can be reached using very little observation data.

11:20 Exploiting Single View Geometry in Pan-Tilt-Zoom Camera Networks

Alberto Del Bimbo (University of Florence, Italy); Fabrizio Dini (University of Florence, Italy); Andrea Grifoni (Thales, Italy); Federico Pernici (University of Florence, Italy)

PTZ (pan-tilt-zoom) camera networks have an important role in surveillance systems. They have the ability to direct the attention to interesting events that occur in the scene. In order to achieve such behavior the cameras in the network use a process known as sensor slaving: one (or more) master camera monitors a wide area and tracks moving targets so as to provide the positional information to one (or more) slave camera. The slave camera foveates at the targets in high resolution. In this paper, we propose a simple method to solve two typical problems that are the basic building blocks to create high level functionality in PTZ camera networks: the computation of the world to image homographies and the computation of camera to camera homographies. The first is used for computing the image sensor observation model in sequential target tracking, the second is used for camera slaving. Finally a cooperative tracking approach exploiting the use of both homographies is presented.

11:40 Collaborative Real-Time Control of Active Cameras in Large-Scale Surveillance Systems

Nils Krahnstoever (General Electric Global Research, USA); Ting Yu (Northwestern University, USA); Ser-Nam Lim (GE Global Research, USA); Kedar Patwardhan (GE Global Research, USA); Peter Tu (GE Global Research, USA)

A comprehensive framework for controlling a set of Pan Tilt Zoom (PTZ) cameras is presented. The goal is to acquire close-up imagery of people in a surveillance site. The control algorithm is driven by the output of a multi-camera, multi-target tracking system that performs person detection and tracking from a set of fixed surveillance cameras. This paper addresses the problem of optimally scheduling the PTZ cameras in real-time under a variety of performance objectives. The scheduler attempts to choose an optimal plan that maximizes the probability of successfully completing a biometrics task. We present a novel objective function that balances the number of captures per target and the quality of captures. We show the results of a system operating in real-time under real-world conditions on four PTZ and four fixed CCTV cameras, all controlled via a single workstation.

12:00 An Efficient Data Driven Algorithm for Multi-Sensor Alignment

Feng Guo (ObjectVideo, USA); Gaurav Aggarwal (University of Maryland, USA); Khurram Shafique (Objectvideo, USA); Xiaochun Cao (ObjectVideo, USA); Zeeshan Rasheed (Objectvideo, USA); Niels Haering (ObjectVideo, USA)

This paper describes how model-specific constraints and domain specific knowledge can be utilized to develop efficient sampling based algorithms for robust model estimation in the presence of outliers. As a special case, a robust algorithm for homography estimation is proposed that exploits the invariance of co-linearity under homography to improve efficiency in noisy scenarios. Unlike most existing approaches, the proposed algorithm does not make any assumption regarding the data distribution, data specific properties or availability of large amount of data. The proposed estimation algorithm is applied for multiple applications involving large sensor networks. These include estimation and maintenance of geo-registration by fusing observations from different modalities, such as RADAR and Automatic Identification System (AIS), and data-driven estimation (using target observations) of the relative topology of cameras with overlapping fields of view. Qualitative and quantitative results are presented that show the ability of the proposed algorithm to handle large fraction of outliers in the data, spatial noise, and high traffic densities, which are defining characteristics of these applications.

Multi-camera networks: applications

16:00 Multi-Camera Multi-Person 3D Space Tracking with MCMC in Surveillance Scenarios

Jian Yao (IDIAP, Switzerland); Jean-Marc Odobez (IDIAP, Switzerland)

We present an algorithm for the tracking of a variable number of 3D persons in a multi-camera setting with partial field-of-view overlap. The multi-object tracking problem is posed in a Bayesian framework and relies on a joint multi-object state space with individual object states defined in the 3D world. The Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) method is used to efficiently search the state-space and recursively estimate the multi-object configuration. The paper presents several contributions: i) the use and extension of several key features for efficient and reliable tracking (e.g. the use of the MCMC framework for multiple camera MOT; the use of powerful human detector outputs in the MCMC proposals to automatically initialize/update object tracks); ii) the definition of appropriate prior on the object state, to take into account the effects of 2D image measurement uncertainties on the 3D object state estimation due to depth effects; iii) a simple rectification method aligning people 3D standing direction with 2D image vertical axis, allowing to obtain better object measurements relying on rectangular boxes and integral images; iv) representing objects with multiple reference color histograms, to account for variability in color measurement due to changes in pose, lighting, and importantly multiple camera view points. Experimental results on challenging real-world tracking sequences and situations demonstrate the efficiency of our approach.

16:20 Multi-camera Matching under Illumination Change Over Time

Bryan Prosser (Queen Mary, University of London, United Kingdom); Shaogang Gong (Queen Mary, Univ. of London, United Kingdom); Tao Xiang (Queen Mary, University of London, United Kingdom)

Illumination differences between disjoint cameras can have a dramatic effect on the appearance of objects, thus increasing the difficulty of multi-camera object association. Although methods to model these inter-camera illumination conditions exist, they often rely on static illumination conditions and are unable to cope with illumination changes over time. In this paper we propose a novel method for multi-camera object association based on adapting a learned inter-camera illumination mapping function to new illumination conditions over time without the need for a manual training stage using new foreground objects. Comparative experiments are carried out using challenging data taken from a disjoint camera network. The results demonstrate that the proposed method outperforms a number of existing methods given changing illumination conditions.

16:40 Video Surveillance using a Multi-Camera Tracking and Fusion System

Zhong Zhang (ObjectVideo, USA); Andrew Scanlon (ObjectVideo Inc, USA); Weihong Yin (ObjectVideo Inc, USA); Li Yu (ObjectVideo Inc, USA); Peter Venetianer (ObjectVideo, USA)

Usage of intelligent video surveillance (IVS) systems is spreading rapidly. These systems are being utilized in a wide range of applications. In most cases, even in multi-camera installations, the video is processed independently in each feed. This paper describes a system that fuses tracking information from multiple cameras, thus vastly expanding its capabilities. The fusion relies on all cameras being calibrated to a site map, while the individual sensors remain largely unchanged. We present a new method to quickly and efficiently calibrate all the cameras to the site map, making the system viable for large scale commercial deployments. The method uses line feature correspondences, which enable easy feature selection and provide a built-in precision metric to improve calibration accuracy.

Poster session

Quasi-Dense Matching between Perspective and Omnidirectional Images

Lingling Lu (Institute of Automation, Chinese Academy of Sciences, P.R. China); Yihong Wu (Chinese Academy of Sciences, P.R. China)

In this paper, we propose a quasi-dense match propagation algorithm for an image pair taken by a perspective camera and an omni-directional camera, where rectification for the omni-directional image is not required. First, a linear transformation is introduced to identify the area containing the corresponding point candidates. Then, a geometric invariant is computed as the constraint for quasi-dense matching. Finally, combining a best-first strategy of Lhuillier and Quan (2002) with this computed geometric invariant, the quasi-dense point correspondences are calculated. The experiments with real data show that the algorithm of this paper has good performance.

Cooperative Surveillance of Multiple Targets using Mutual Information

Eric Sommerlade (University of Oxford, United Kingdom); Ian Reid (University of Oxford, United Kingdom)

This work presents a method to control multiple, but diverse pan-tilt-zoom cameras which are sharing overlapping views of the same spatial location for the purpose of observation of this scene. We cast this control input selection problem into a information-theoretic framework, where we maximise the expected mutual information gain in the scene model with respect to the observation parameters. The scene model yielding this information comprises several dynamic targets, augmented by one which has not yet been detected. The information content of the former is supplied directly by the uncertainties computed using a Sequential Kalman Filter tracker for the observed targets, while the undetected is modelled using a Poisson process for every element of a common ground plane. Together these yield an information-theoretic utility for each parameter setting for each camera, triggering collaborative explorative behaviour of the system.

Multiple Persons Tracking with Data Fusion of Multiple Cameras and Sensing Floor Using Particle Filters

Taketoshi Mori (The University of Tokyo, Japan); Takashi Matsumoto (The University of Tokyo, Japan); Masamichi Shimosaka (The University of Tokyo, Japan); Hiroshi Noguchi (The University of Tokyo, Japan); Tomomasa Sato (The University of Tokyo, Japan)

Successful multi-target tracking requires locating the targets and labeling their identities. For the multi-target tracking systems, the latter becomes more challenging when the targets frequently interact with each other. In this paper, we propose a method for multiple persons tracking using multiple cameras and sensing floor. Our method estimates 3D positions of human body and head, and labels their identities. Our method is composed of multiple particle filters that interact only in the exclusion occlusion model. Each particle filter tracks each person correctly by integrating information from sensing floor and the target-specific information from multiple cameras. Integration of these two sensors enables complement of each weak point and the correct tracking of the target. Moreover, we develop a new particle filter framework that tracks the human head by using the estimated human body position simultaneously. Our experimental results demonstrate the effectiveness and robustness of the method against several complicated movements of multiple persons. The results also demonstrate that this method can maintain correct tracking when the targets are in close proximity.

A Comparative Error Analysis of Audio-Visual Source Localization

Damien Kelly (Trinity College Dublin, Ireland); Francois Pitie (Trinity College Dublin, Ireland); Anil Kokaram (Trinity College Dublin, Ireland); Frank Boland (Trinity College Dublin, Ireland)

This paper examines the accuracy of audio-video based localization using multiple cameras and multi-microphones. Covariance mapping theory is used to determine the accuracy of audio and video based localization. Both modalities are compared in terms of their ability to provide accurate location estimates of a moving audio-visual source. Relatively, video is found to be significantly more accurate than audio. The problem of audio-video fusion is also examined. The fusion of audio and video location estimates is applied in the audio domain, the video domain and the positional domain. The accuracy of these three fusion strategies for 3D localization are examined from a theoretical basis. The best localization performance is found when fusion is applied in the positional domain. Fusing audio and video data in the video domain is found to exhibit the worst localization performance. This analysis is confirmed by measuring the accuracy of each fusion strategy in localizing a moving audio-visual source.

Towards Audio-Visual On-line Diarization Of Participants In Group Meetings

Hayley Hung (IDIAP Research Institute, Switzerland); Gerald Friedland (International Computer Science Institute, USA)

We propose a fully automated, unsupervised, and non-intrusive method of identifying the current speaker audio-visually in a group conversation. This is achieved without specialized hardware, user interaction, or prior assignment of microphones to participants. Speakers are identified acoustically using a novel on-line speaker diarization approach. The output is then used to find the corresponding person in a four-camera video stream by approximating individual activity with computationally efficient features. We present results showing the robustness of the association on over 4.5 hours of non-scripted audio-visual meeting data.

Finding Speaker Face Region by Audiovisual Correlation

Yuyu Liu (The University of Tokyo, Japan); Yoichi Sato (Tokyo University, Japan)

The ability to find the speaker face region in a video is important in various application areas. In this work, we develop a novel technique to find this region robustly against different views and complex backgrounds using gray images only. The main thrust of this technique is to integrate audiovisual correlation analysis into an image segmentation framework to extract the speaker face region. We first analyze the video in a time window and evaluate the audiovisual correlation locally at each pixel position using a novel statistical measure based on Quadratic Mutual Information. As only local visual information is adopted in this stage, the analysis is robust against the view change of the human face. Analyzed correlation is then incorporated into Graph Cut-based image segmentation, which optimizes an energy function defined over multiple video frames. As this process can find the global optimum segmentation with image information balanced, we thus can extract a reliable region aligned to real visual boundaries. Experimental results demonstrate the effectiveness and robustness of our method.

State-Based Visibility for 3D Reconstruction from Multiple Views

Liuxin Zhang (Beijing Institute of Technology, P.R. China); Yumo Yang (Beijing Institute of Technology, P.R. China); Yunde Jia (Beijing Institute of Technology, P.R. China)

Estimating visibility is one of the most important ingredients for any multi-view reconstruction using volumetric scene representation. In this paper, we propose a simple approach to estimating visibility based on current state of the scene, which is implicitly represented as the zero level set of a function. A one-pass algorithm is adopted to determine the regions of a 3D space visible to several given viewpoints efficiently. A new variational model is introduced based on the proposed state-based visibility method for multi-view reconstruction. We cast the reconstruction problem as an optimization of a novel energy functional amenable for minimization with an Euler-Lagrange driven evolution. The proposed algorithm has been applied to both synthetic and real datasets with promising results.

Estimation of a 3D motion field from a multi-camera array using a multiresolution Gaussian mixture model

Roland Wilson (Warwick, United Kingdom); Adam Bowen (University of Warwick, United Kingdom); Andrew Mullins (University of Warwick, United Kingdom); Nasir Rajpoot (University of Warwick, United Kingdom)

The problem of modelling geometry for video based rendering has been much studied in recent years, due to the growing interest in `free viewpoint' video and similar applications. Common approaches fall into two categories: those which approximate surfaces from dense depth maps obtained by generalisations of stereopsis and those which employ an explicit geometric representation such as a mesh. While the former have generality with respect to geometry, they are limited in terms of viewpoint; the latter, on the other hand, sacrifice generality of geometry for freedom to pick an arbitrary viewpoint. The purpose of the work reported here is to bridge this gap in object representation, by employing a stochastic model of object structure: a multiresolution Gaussian mixture. Estimation of the model and tracking it through time from multiple cameras is achieved by a multiresolution stochastic simulation. After a brief outline of the method, its use in modelling human motion using data from local and other sources is presented to illustrate its effectiveness compared to the current state of the art.

A Stochastic Quality Metric for Optimal Control of Active Camera Network Configurations for 3D Computer Vision Tasks

Adrian Ilie (University of North Carolina at Chapel Hill, USA); Greg Welch (University of North Carolina, USA); Marc Macenko (University of North Carolina at Chapel Hill, USA)

We present a stochastic state-space quality metric for use in controlling active camera networks aimed at 3D vision tasks such as surveillance, motion tracking, and 3D shape/appearance reconstruction. Specifically, the metric provides an estimate of the aggregate steady-state uncertainty of the 3D resolution of the objects of interest, as a function of camera parameters such as pan, tilt, and zoom. The use of stochastic state-space models for the quality metric results in the ability to model and accommodate virtually all traditional quality factors, such as visibility, field of view, occlusion, resolution, surface normals, image contrast, focus, and depth of field. In addition, the stochastic state-space approach naturally addresses camera networks that are aided by other sensing modalities. We begin by surveying the traditional quality factors. We then present our new quality metric, aided by some background in the relevant stochastic state-space models, and an evaluation strategy that scales the computation of our metric to allow its use in a real-time active camera network system. Finally we present some simulation results that illustrate the incorporation of some traditional quality factors, and the use of our metric and evaluation strategy for some simulated scenes containing multiple objects of interest.

Largest Silhouette-Equivalent Volume for 3D Shapes Modeling without Ghost Object

Brice Michoud; Saida Bouakaz; Erwan Guillou; Hector Briceno (LIRIS - CNRS - Université Lyon 1, France)

In this paper, we investigate a practical framework to compute a 3D shape estimation of multiple objects in real-time from silhouettes in multi-view environments. A popular method called Shape From Silhouette (SFS), computes a 3D shape estimation from binary silhouette masks. This method has several limitations: The acquisition space is limited to the intersection of the camera viewing frusta ; SFS methods reconstruct some ghost objects which do not contain real objects, especially when there are multiple real objects in the scene. In this paper we propose two contributions to overcome these limitations. First, using a new formulation of SFS approach, our system reconstructs objects with no constraints on camera placement and their visibility. Second, a new theoretical approach identifies and removes ghost objects. The reconstructed shapes are more accurate than current silhouette-based approaches. Reconstructed parts are guaranteed to contain real objects. Finally, we present a real-time system that captures multiple and complex objects moving through many camera frusta to demonstrate the application and robustness of our method.

Multiple Camera Person Tracking in multiple layers combining 2D and 3D information

Dejan Arsic (Technische Universität München, Germany); Bjoern Schuller (Technische Universitaet Muenchen, Germany); Gerhard Rigoll (Technische Universitaet Muenchen, Germany)

CCTV systems have been introduced in most public spaces in order to increase security. Video outputs are observed by human operators if possible but mostly used as a forensic tool. Therefore it seems desirable to automate video surveillance systems, in order to be able to detect potentially dangerous situations as soon as possible. Multi camera systems have seem to be the prerequisite for huge spaces where frequently occlusions appear. In this treatise we will present a system which robustly detects and tracks objects in a multi camera environment and performs a subsequent behavioral analysis based on luggage related events.

3D Markerless human limb localization through robust energy minimization

Marco Marcon (Politecnico di Milano, Italy); Massimiliano Pierobon (Politecnico di Milano, Italy); Augusto Sarti (Politecnico di Milano, Italy); Stefano Tubaro (Politecnico di Milano, Italy)

Markerless human tracking addresses the problem of estimating human body motion in non-cooperative environments. Computer Vision techniques combined with Pattern Recognition theory serve the purpose of extracting information on human body postures from video-sequences, without the need of wearable markers. Multi-camera systems further enhance this kind of application providing frames from multiple viewpoints. This work tackles the application of multi-camera posture estimation through the use of a multi-camera environment, also known as "smart space". A 3D skeleton structure and geometrical descriptors of human muscles are fitted to the volumetric data to directly recover 3D information. 3D skeleton deformations and bio-mechanical constraints on joint models are used to provide posture information at each frame. The proposed system does not require any pre-initialization phase and automatically adapt the skeleton and the volumetric occupation of each limb to the actor physiognomy independently from the pose. Exhaustive tests were performed to validate our approach.

Camera-Clustering for Multi-Resolution 3-D Surface Reconstruction

Andrei Zaharescu (INRIA Rhone-Alpes, France); Cedric Cagniart (TU Berlin, Germany); Slobodan Ilic (TU Berlin / DT Laboratories, Germany); Radu Horaud (INRIA Rhone-Alpes, France); Edmond Boyer (INRIA Rhone-Alpes, France)

In this paper we propose a framework for piecewise mesh-based 3D reconstruction from a set of calibrated images. Most of the available approaches are global and require the use of all available images at once. Instead, we use subsets of images and evolve parts of the surface corresponding to those images. Our main contribution is an approach to partitioning of the camera images, which can be either semi-automatic, trough clustering, or user guided, trough a geometric modeling interface. The sub-parts of the surface corresponding to camera subsets are independently evolved at multiple mesh resolutions. This allows to handle large scenes and to increase the mesh resolution in surface parts containing high levels of detail at reduced memory and computational costs. We demonstrate the versatility of our approach on different data sets and with different camera deployment. Finally, comparing the piecewise and global reconstructions with groundtruth, we find no significant loss in the overall reconstruction quality.

A multi-sensor approach for People Fall Detection in home environment

Cosimo Distante (CNR, Italy); Alessandro Leone (CNR, Italy); Piero Malcovati (University of Pavia, Italy)

This paper presents a hardware and software framework for reliable fall detection in the home environment, with particular focus on the protection and assistance to the elderly. The integrated prototype includes three different sensors: a 3D Time-Of-Flight range camera, a wearable MEMS accelerometer and a microphone. These devices are connected with custom interface circuits to a central PC that collects and processes the information with a multi-threading approach. For each of the three sensors, an optimized algorithm for fall-detection has been developed and benchmarked on a collected mulitimodal database. This work is expected to lead to a multi-sensory approach employing appropriate fusion techniques aiming to improve system efficiency and reliability.

Eyewear Selector

Oscar Déniz-Suárez (University of Castilla-La Mancha, Spain); Modesto Castrillón-Santana (University of Las Palmas de Gran Canaria, Spain); Javier Lorenzo-Navarro (Univ. de Las Palmas de Gran Canaria, Spain); Mario Hernandez Tejera (Universidad de Las Palmas de Gran Canaria, Spain); Luis Anton Canalis (Universidad de Las Palmas de Gran Canaria, Spain); Gloria Bueno Garcia (Universidad de Castilla-La Mancha, Spain)

The widespread availability of portable computing power and inexpensive digital cameras are opening up new possibilities for retailers. One example is in optical shops, where a number of systems exist that facilitate eyeglasses selection. These systems are now more necessary as the market is saturated with an increasingly complex array of lenses, frames, coatings, tints, photochromic and polarizing treatments, etc. Research challenges encompass Computer Vision, Multimedia and Human-Computer Interaction. Cost factors are also of importance for widespread product acceptance. This paper describes a low-cost system that allows the user to visualize different glass models in live video. The user can also move the glasses to adjust its position on the face. Experiments show the potential of the system.

A Noise-Aware Filter for Real-Time Depth Upsampling

Derek Chan (Stanford University, USA); Hylke Buisman (Stanford University, USA); Christian Theobalt (Stanford University, USA); Sebastian Thrun (Stanford University, USA)

A new generation of active 3D range sensors, such as time-of-flight cameras, enables recording of full-frame depth maps at video frame rate. Unfortunately, the captured data are typically starkly contaminated by noise and the sensors feature only a rather limited image resolution. We therefore present a pipeline to enhance the quality and increase the spatial resolution of range data in real-time by upsampling the range information with the data from a high resolution video camera. Our algorithm is an adaptive multi-lateral upsampling filter that takes into account the inherent noisy nature of real-time depth data. Thus, we can greatly improve reconstruction quality, boost the resolution of the data to that of the video sensor, and prevent unwanted artifacts like texture copy into geometry. Our technique has been crafted to achieve improvement in depth map quality while maintaining high computational efficiency for a real-time application. By implementing our approach on the GPU, the creation of a real-time 3D camera with video camera resolution is feasible.

Fusion of Time of Flight Camera Point Clouds

James Mure-Dubois (University of Neuchatel, Switzerland); Heinz Hügli (University of Neuchâtel, Switzerland)

Recent time of flight cameras deliver range images (2.5D) in real-time, and can be considered as a significant improvement when compared to conventional (2D) cameras. However, the range map produced has only a limited extent, and suffers from occlusions. In this paper, we investigate fusion methods for partially overlapping range images, aiming to address the issues of lateral field of view extension (by combining depth images with parallel view axes) and occlusion removal (by imaging the same scene from different viewpoints).

Shape from Probability Maps with Image-Adapted Voxelization

Jordi Salvador Marcos (Technical University of Catalonia (UPC), Spain); Josep Casas (UPC - Technical University of Catalonia, Spain)

This paper presents a Bayesian framework for Visual Hull reconstruction from multiple camera views with a 3D sampling scheme based on an irregular 3D grid, which becomes regular once projected onto the available views. The probabilistic framework consists in establishing a foreground probability for each pixel in each view rather than segmenting in order to obtain binary silhouettes of the foreground elements. Next, a Bayesian consistency test labels the occupancy of each image-adapted 3D sample. The proposed method, using image-adapted 3D sampling in the Bayesian framework, is compared to a shape-from-silhouette implementation with image-adapted voxelization, where the input data are binary silhouettes instead of probability maps; we also compare its performance to a state-of-the-art method based on regular 3D sampling with binary silhouettes and SPOT projection test.

Colour Constancy Techniques for Re-Recognition of Pedestrians from Multiple Surveillance Cameras

Alberto Colombo (Kingston University, United Kingdom); James Orwell (Kingston University, United Kingdom); Sergio Velastin (Kingston University, United Kingdom)

This paper presents work towards a system for tracking the movements of a pedestrian as they move between the multiple sensors comprising a surveillance system. The colour appearance of the observations is an important cue: it is useful to achieve good color constancy between the colour values associated with each camera. A novel method for estimating the appropriate transform between each camera's colour space is proposed, using covariance of the foreground data collected from each camera. Simulations are used to demonstrate that the method only works if the covariance has a sufficiently high ratio between its eigenvalues. The covariance matrices for foreground data collected from 29 surveillance cameras are estimated and shown to have a sufficiently high ratio. The discriminative power of colour-based appearance descriptors is evaluated using several types of colour constancy methods. The proposed method leads to a significant improvement in the simplest and best performing (mean) colour descriptor. It is shown how these descriptors can be integrated into a probabilistic framework for tracking pedestrians from multiple surveillance cameras.

Map-based Active Leader-Follower Surveillance System

Himaanshu Gupta (ObjectVideo, USA); Xiaochun Cao (ObjectVideo, USA); Niels Haering (ObjectVideo, USA)

We propose a generic framework for an active leader-follower surveillance system. The system can ingest inputs from a variety of multi-modal leader sensors: Radars, Automatic Identification Systems (AIS), or cameras. Rule and learning based pattern recognition techniques are performed on the leader sensors to infer unusual behaviors and events. In order to calibrate the follower PTZ cameras to the leader sensors, we break up the calibration process into two independent steps: mapping leader cameras' image coordinates to latitude/longitude values, and mapping the latitude/longitude values to the follower's pan-tilt-zoom settings. This calibration method provides straightforward scalability to complex camera networks since newly added cameras need to be calibrated only with respect to the site map (latitude/longitude), and the Radar or AIS leader sensors also fit in seamlessly. Upon detection of an unusual event by one of the leader sensors, follower PTZ cameras move to the location where the event was triggered, detect the target of interest in the field of view, and then track the target actively by automatically adjusting the PTZ settings. We also propose a system to learn various PTZ camera characteristics which facilitate stable and accurate control of the camera during active tracking. Finally, we evaluate the performance of each of the components individually, as well as that of the entire system.

Multi-modal Video Surveillance aided by Pyroelectric Infrared Sensors

Michele Magno (University of Bologna, Italy); Federico Tombari (University of Bologna, Italy); Davide Brunelli (University of Bologna, Italy); Luigi Di Stefano (Universita' di Bologna, Italy); Luca Benini (Unversity of Bologna, Italy)

The interest in low-cost and small size video surveillance sys- tems able to collaborate in a network has been increasing over the last years. Thanks to the progress in low-power design, research has greatly reduced the size and the power consumption of such distributed embed- ded systems providing flexibility, quickly deployment and allowing the implementation of effective vision algorithms performing image process- ing directly on the embedded node. In this paper we present a multi-modal video sensor node designed for low-power and low-cost video surveillance able to detect changes in the environment. The system is equipped with a CMOS video camera and Pyroelectric InfraRed (PIR) sensors exploited to reduce remarkably the power consumption of the system in absence of events. The on-board microprocessor implements a NCC algorithm. We analyze different con- figurations and characterize the system in term of runtime execution and power consumption.

Object Detection and Matching with Mobile Cameras Collaborating with Fixed Cameras

Alexandre Alahi (EPFL, Switzerland); Michel Bierlaire (EPFL, Switzerland); Murat Kunt (EPFL Lausanne, Switzerland)

A system is presented to detect and match objects with mobile cameras collaborating with fixed cameras observing the same scene. No training data is needed. Various object descriptors are studied based on grids of region descriptors. Region descriptors such as histograms of oriented gradients and covariance matrices of different set of features are evaluated. A detection and matching approach is presented based on a cascade of descriptors outperforming previous approaches. The object descriptor is robust to any changes in illuminations, viewpoints, color distributions and image quality. Objects with partial occlusion are also detected. The dynamic of the system is taken into consideration to better detect moving objects. Qualitative and quantitative results are presented in indoor and outdoor urban scenes.

Video alignment for difference spotting

Ferran Diego (Universitat Autònoma de Barcelona, Spain); Daniel Ponsa (UAB, Spain); Joan Serrat (Universitat Politècnica de Catalunya, Spain); Antonio Lopez (UAB, Spain)

We address the synchronization of a pair of videos sequences captured from moving vehicles and the spatial registration of all the temporally corresponding frames. The final goal is to fuse the two videos pixel--wise and compute their pointwise differences. Video synchronization has been attempted before but often assuming restrictive constraints like fixed or rigidly attached cameras, simultaneous acquisition, known scene point trajectories etc. which to some extent limit its practical applicability. We intend to solve the more difficult problem of independently moving cameras which follow a similar trajectory, based only on the fusion of image intensity and GPS data information. The novelty of our approach is the probabilistic formulation and the combination of observations from these two sensors, which have revealed complementary. Results are presented in the context of vehicle pre--detection for driver assistance, on different road types and lighting contiditions.

Performance Evaluation of Multisensor Architectures for Tracking

Stefano Maludrottu (University of Genova, Italy); Alessio Dore (University of Genova, Italy); Hany Sallam (University of Genova, Italy); Carlo Regazzoni (University Of Genova, Italy)

Performance evaluation of data fusion systems for tracking is a key task for the correct design of effective and robust architecture. In fact in many applications, as videosurveillance or radar flight tracking, a precise target localization is required over time. In these domains multisensor approaches are successfully employed because of the capability of resolving problems caused by misleading observations with the redundancy introduced by the possibility of having multiple measurements for the same target. The architecture of these systems is typically complex and composed by many sensors to monitor wide areas and fusion modules that provide a unique scene representation. Then several processing units and communications links are involved in this process and their behavior affects the performances. In this paper a model of a data fusion system for tracking is proposed that takes into account different aspects of such system architectures to assess their performances for what concerns the algorithms accuracy and also to consider communication and computational complexity issues that can arise in complex multisensor systems.

A New Approach for Target Motion Analysis in a Binary Sensor Network

Adrien Ickowicz (IRISA, France); Jean-Pierre Le Cadre (IRISA / CNRS Rennes, France)

The aim of this paper is to present a new concept for target motion analysis within a binary sensor network. For the sake of sim- plicity, we focus on a constant target motion. The binary information represents a very rough information about the perception of the target motion by an elementary sensor, i.e. is the target approaching or going away. Collecting these binary informations, a first step is to determine the information we can extract at the network level about target motion. Then, based on this step, new concepts are introduced for infering the target motion parameters. One is based upon the separation properties and relies on the SVM formalism; while the other one uses the concept of the velocity plane and the PPR (Projection Pursuit Regression) frame- work. Moreover, theoretical results about the convergence of this method are also presented.

Efficient Resource Allocation using a Multiobjective Utility Optimisation Method

Stephan Matzka (Heriot-Watt University, United Kingdom); Yves Petillot (HWU, United Kingdom); Andrew Wallace (Heriot-Watt University, United Kingdom)

In this paper we present an extension for two recent active vision systems proposed in Navalpakkam and Itti [1], and in Frintrop [2]. The novelty of our proposed system is twofold: first it extends the existing approaches using both prior and dynamic contextual knowledge, enabling to adapt the proposed system to the present environment. Second, the decision making process intuitively used in [1,2] is formalised in this paper and put into the context of multiobjective optimisation using a utility concept. We discuss three different saliency algorithms to be used in the system as well as three different methods to determine common utility. Our presented system is quantitatively evaluated using a motorway traffic sequence recorded by a test vehicle equipped with a multimodal sensor system.

On the Benefits of Using Gyroscope Measurements with Structure from Motion

Adel Fakih (Univeristy of Waterloo, Canada); John Zelek (University of Waterloo, Canada)

This paper is concerned with the benefits of using rotational measurements of a gyroscope in conjunction with optical flow to determine the instantaneous rigid motion. It presents an experimental study focusing on three main aspects: speedup, resolving ambiguities due to local minima and accuracy. For this sake, we developed a method to minimize simultaneously the reprojected optical flow error and the deviation from the gyro measurements, in line with the optimal Structure from Motion techniques. We show that the gyro measurements can be used in a procedure to initialize the iterative estimation which results in a considerable speedup. We show also that minimizing both the deviation from the measured flow and the gyro measurements gives better results for the rotational velocity when the gyro measurements are not very noisy. It doesn't affect the accuracy of the translational velocity, however if the gyro measurements are very noisy it might lead to erroneous translational estimates.

A Monte Carlo Based Framework for Multi-Target Detection and Tracking Over Multi-Camera Surveillance System

Chingchun Huang (National Chiao Tung University, Taiwan); Sheng-Jyh Wang (National Chiao Tung University, Taiwan)

In the paper, we proposed a system for automatic detection and tracking of multiple targets in a multi-camera surveillance zone. In each camera view of this system, we only need a simple object detection algorithm, such as background subtraction. The detection results from multiple cameras are fused into a posterior distribution, named TDP, based on the Bayesian rule. This TDP distribution indicates the likelihood of having some moving elements on the ground plane. To properly handle the tracking of multiple moving targets over time, a sample-based framework, which combines Markov Chain Monte Carlo (MCMC), Sequential Monte Carlo (SMC), and Mean-Shift clustering, is proposed. The MCMC is used to handle the occurrence of new targets. The SMC is used to track existing targets over time. The Mean-Shift clustering is adopted to automatically identify new comers. With the Monte Carlo based framework, the detection and tracking of multiple targets can be achieved in a unified and seamless manner. The detection and tracking accuracy is evaluated by both synthesized videos and real videos. The experimental results show that the proposed system can successfully track a varying number of people accurately.

Multi-Camera Visual Surveillance for Motion Detection, Occlusion Handling, Tracking and Event Recognition

Oytun Akman (Delft University of Technology, The Netherlands); Aydin Alatan (Middle East Technical University, Turkey); Tolga Ciloğlu (Middle East Technical University, Turkey)

In this paper, we propose novel methods for background modeling, occlusion handling and event recognition by using multi-camera configurations. Homography-related positions are utilized to construct a mixture of multivariate Gaussians to generate a background model for each pixel of the reference camera. Occlusion handling is achieved by generation of the top-view via trifocal tensors, as a result of matching over-segmented regions instead of pixels. The resulting graph is segmented into objects after determining the minimum spanning tree of this graph. Tracking of multi-view data is obtained by utilizing measurements across the views in case of occlusions. Finally, the resulting trajectories are classified by GM-HMMs, yielding better results for using together all different view trajectories of the same object. Hence, multi-camera sensing is fully exploited from motion detection to event modeling.

Composite Spatio-Temporal Event Detection in Multi-Camera Surveillance Networks

Yun Zhai (IBM Watson Research Center, USA); Ying-li Tian (IBM T.J. Watson Research Center, USA); Arun Hampapur (IBM T.J. Watson, USA)

In this paper, we present a composite event detection system for multi-camera surveillance networks. The proposed framework is able to handle correlations between primitive events that are generated from either a single camera view or multiple camera views with spatial and temporal variations. Composite events are represented in the form of full binary tree, where the leaves nodes represent the primitive events, the root node represents the target composite event, and the middle node represent the defined rules. The multi-layer design of the composite events provides a great extensibility and flexibility to users with different applications. The high-level composite events are represented by a standardized XML-style description language, which is used for inter-agent communications and event detection module construction. A set of graphical user interfaces are developed for conveniently defining both primitive and high-level composite events. The proposed system is designed in the distributed form, where different components of the system can be deployed on separate work locations and communicate with each other over the network. The capabilities and effectiveness of the proposed composite event detection system have been demonstrated in several real-life applications.