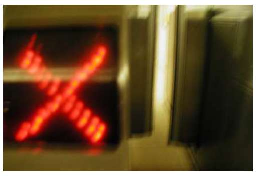

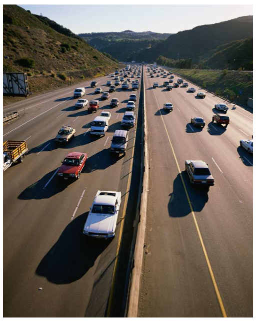

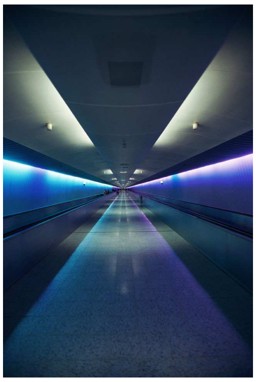

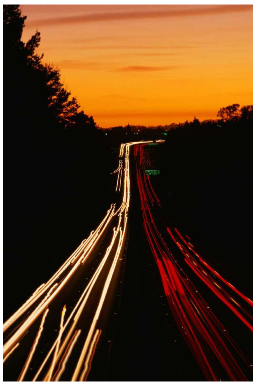

Pictures: DLR

Workshop
Sensor- und Datenfusion – Architekturen und Algorithmen

20. November, 2009

10:00 Uhr – 16:00 Uhr
Newton Kabinett
Rudower Chaussee 17
12489 Berlin, Germany

Organisation:

Humboldt-Universität zu Berlin
Institut für Informatik, Computer Vision

Mit DIN, OpTecBB (Branchentransferstelle Optik), WISTA MANAGMENT GMBH, DLR-TS

Die Fusion von Daten unterschiedlicher digitaler Sensoren (Kameras, Laserscanner, Hyperspektralscanner und Radarsysteme) erhöht die Auflösung, die Genauigkeit und die Zuverlässigkeit des Datenprodukts. Darüber hinaus können den Anwendern völlig neue Datenprodukte angeboten werden.

Jedoch ist festzustellen, dass die Fusion für Anwendungen, zum Beispiel in Photogrammetrie und Fernerkundung, bisher lediglich auf dem untersten Niveau – dem Datenlevel – teilweise auch als kommerzielle Leistung durchgeführt wird. Damit sind aber die Möglichkeiten und Implementierungen von Sensor- und Datenfusion nicht ausgeschöpft. Es ist zu erwarten, dass sie sich in den nächsten Jahren drastisch ausweiten werden.

Innovativ ist die notwendige Betrachtung der verschiedenen Fusionsebenen (Multi-Level-Sensorfusion). Neben den Anwendungen in Photogrammetrie und Fernerkundung sollen auch Applikationen und Anwendungen aus benachbarten Fachgebieten zur Navigation und Orientierung, zum Tracking und zu Fahrerassistenzsystemen berücksichtigt werden. Aus der Analyse von Anwendungen, Algorithmen und Architekturen von Datenfusionsverfahren ist es möglich, entsprechende Gütekriterien abzuleiten.

In diesem Workshop sollen Beiträge zu den wissenschaftlichen Grundlagen für die Vergleichbarkeit der Fusionsprodukte geschaffen werden. Sie sollen Grundlage für eine Normung der aus Sensor- und Datenfusion abgeleiteten Produkte geben, um den Anwendern auf dem globalen Markt die Kriterien zur Entscheidungsfindung für Systembeschaffungen sowie Qualitätsstandards für ein Endprodukt geben zu können.

- Allgemeine Datenanalyse und Möglichkeiten für die Fusion
    - hochauflösende digitale Kameras
    - Laserscanner
    - multispektrale- und hyperspektrale Systeme
    - Radar
- Untersuchung von Kombinationen (Multi-Level-Sensorfusion)
    - Realisierungsweg für die Sensorfusion,
    - Güte / Zuverlässigkeit und Auswerteverfahren für verfügbare Sensordaten
    - Einbeziehung von verschiedenen Ebenen von Sensorfusion (Multi-Level-Sensorfusion)
    - Gemeinsame Verarbeitung der verschiedenen Fusions-Ebenen
- Analyse der Synergieeffekte
    - Fusion-Feedback
    - Hervorhebung eines Aufmerksamkeitsbereichs (Area of Interests)
    - Lösung von widersprüchlichen Angaben

Weitere Informationen + Teilnahmezusage:
http://tinyurl.com/FusionDE
Email: **reulke@informatik.hu-berlin.de** oder **m.misgaiski@gmx.de**

Agenda zum Workshop
Sensor- und Datenfusion – Architekturen und Algorithmen
20. November, 2009
Newton Kabinett
Rudower Chaussee 17
12489 Berlin, Germany

10:00 - 10:15 **Begrüßung & Einführung**
Prof. Dr. Ralf Reulke
Humboldt-Universität zu Berlin
Institut für Informatik, Computer Vision;
Deutsches Zentrum für Luft- und Raumfahrt
Institut für Verkehrssystemtechnik

10:15 - 10:45 **Sensorfusion as key technology for future driver assistance and safety systems**
Dr. J. Dickmann
Daimler AG

10:45 - 11:15 **Multi-Level Fusion for Cooperative Cognitive Automobiles**
Prof. Dr. C. Stiller
Institut für Mess- und Regelungstechnik
Karlsruhe Institute of Technology (KIT)

11:15 - 11:45 **Data fusion for optical navigation systems**
Dr. A. Börner
Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Optische Informationssysteme
Informationsverarbeitung optischer Systeme

11:45 - 13:30 **Mittagspause**

13:30 - 14:00 **Multicamera, distributed and heterogeneous sensor systems: Architectures and algorithms for people surveillance**
Prof. R. Cucchiara
Università di Modena e Reggio Emilia

14:00 - 14:30 **Swarming Machines & Sensor Fusion**
What can we learn from biology?
Prof. Dr. V. V. Hafner
Humboldt-Universität zu Berlin
Institut für Informatik

14:30 - 15:00 ***(Titel wird sobald als möglich nachgereicht)***
Prof. Dr. R. Rojas
Freie Universität Berlin
Department of Mathematics and Computer Science

15:00 - 16:00 **Diskussion mit anschließendem Get-Together**

Sensorfusion as key technology for future driver assistance and safety systems

Dr. **J. Dickmann**, N. Appenrodt, Dr. O. Löhlein, Dr. M. Mekhaiel, Dr. M. Mählisch, M. Muntzinger, Dr. R. Schweiger, S. Hahn, DAIMLER AG, 89081-Ulm, Germany

Abstract

The capability to assess and perceive the actual driving situation in complex traffic situations is the key enabler for future vehicle comfort- and safety systems. The symbiotic exploitation of the electromagnetic spectrum by means of Radar- and optical sensors like Scanner and Vision sensors allows the comprehensive and precise detection even at adverse conditions. On the other hand safety has to be affordable. Hence, the question for the optimum compromise between required system performance and minimum possible costs become more and more important. The decision for the sensor set-up and the way how to perform the sensor data fusion as well as the interface definition between OEM and supplier is one part of the answer. The article decribes one possible fusion approach.

1. Introduction

Since the introduction of simple driver assistant functions like ultra sonic based park assist systems, sensor fusion has been used to achieve a market relevant performance to cost ratio. The key issue was, to exploit the information of very simple (and cost effective) sensors in an intelligent way to enable the required function. This classical role as function enabler continues with the complexity of task to be performed.

With the introduction of the first brake assist function and it´s logical next step the emergency braking system (e.g. PRE-SAFE Brake in the Mercedes-S-Class) a combination of near- and far range radars was fused to provide the required environmental information. Up to now, assistance functions on the market concentrate on the use of one single sensor technology like ultra-sonic-, radar- fixed beam lidar-sensors and mono-vision. More sophisticated optical sensors like scanning lidar and stereo-vision sensors have made a great stride ahead to meet vehicle relevant maturity and packaging constraints, which will make employment in vehicle systems very likely in the near future.

Sensor fusion enhances the information of one single sensor technology, as e.g. long range radar and near range radars for collision mitigation, to a level which enables the driving tasks.

This is sufficient for functions which operate in clear driving situations e.g. highways and/or have to perform moderate actions like distance control.

Future comfort- and especially safety functions will more and more address urban regions with dense traffic and therefore have to perform more complex tasks in more complex traffic environments. Thus, the near and mid range distance of the vehicles environment will become more important along with a wider lateral observation horizon in order to cover e.g. crossing scenarios or classical pre-crash situations. This imposes challenging requirements for the environmental sensing, since it translates into dramatically shrinking time scales in terms of observation horizons and reaction times compared to classical ACC and todays collision mitigation functions.

Hence, a much faster up-date rate combined with more detailed and precise information about the traffic environment is mandatory to allow for a reliable situation assessment, especially for future safety systems. This can be achieved over a two way strategy. First, enhance the sensing performance of the environmental sensors in terms of higher spatial resolution towards wide field of view image like properties and add classification knowledge as one perception aid to the sensor information. Second, synergetic exploitation of the electromagnetic spectrum by fusion of different physical sensor technologies.

Since safety and comfort functions have to be affordable, it is obvious, that solutions on the sensor side as well as for the fusion-approach have to be developed and identified that enable the best performance at automotive market relevant cost structures.

2. Sensors

Radar sensors have long been the leading edge in vehicular remote sensing. They determine highly precise distance information and provide instantaneously the corresponding target/object - velocity at nearly all environmental conditions. State of the art radars operate in multi mode covering long and short range distances in one sensor. One field for optimization of todays radars is the limited field of view along with limited angular resolution, which limits the precision of the target/object localization. Up to now radars suffer from limited classification capability. On research level, imaging radar approaches are investigated, which definitely will close these present performance gaps to make radar an utmost device [1].

Recent advances in scanning lidar technology make it an interesting candidate for remote sensing, which offers excellent spatial information along with an extremely wide field of view starting from ±45° up to ±120° and cover ranges from 0 up to 200 m. Recent filter development on the supplier side has enabled velocity information to be provided nearly simultaneously to the spatial information. Since lidar is an optical sensor, it suffers like vision systems from a limited all weather capability. However, the high end lidar versions e.g. from Hella KGaA Hueck & Co, have considerably improved even in this area. The high information density allows much better object localisation including dimension and orientation information of the objects. As long as no classification information is required, they could operate as the ideal tandem arrangement to a radar. A very prominent example is described in section 5 for crossing scenarios where a very fast wide field of view and very precise object localisation is required to allow for pre-crash detection.

Vision systems add those information, which cannot easiliy or even never be provided by radar or lidar systems. Such as classification information lane-prediction, traffic sign recognition or classification of pedestrians, cylist, vehicle-type etc. [2]. In appropriate constellations, sophisticate image processing for stereo vision allows for velocity information making vision a powerful sensor [3]. Infrared cameras extend the use availability into the night. Thus, depending on the safety or comfort function to be realized, vision is the necessary supplement to both lidar and radar.

3. Fusion Concept: One for all or unity in diversity?

Sensor fusion allows the exploitation of the best from all sensors to overcome the shortcomings of one individual sensor to form an enriched knowledge about the environment as performed with only one sensor or one sensor technology. Also here, for an OEM the task is to find the optimum compromise between performance and cost.

The present day standard approach of sensor fusion in the automotive industry is high level or also called object level fusion. Each sensor provides high level track information which is aligned in space and on temporal base, before they are fused in a track level fusion module. From an economical standpoint, the advantages so far are manifold. Among others, there is flexibility. On object level, it is easy to modify the present constellation of the sensor setup or later on incorporation of innovative new sensor types. High tolerance against last minute changes in one sensor. The data rate is moderate to allow for automotive bus-standards like CAN-bus, and so on. The interface between supplier and OEM is defined at the object level. However, a potential drawback is the application specific definition of objects, which leads to some degree of extra effort in adaptation from one application to another and a potential loss in raw data information during the object formation process. Another potential drawback is a potential lack in competition advantage since each OEM may also get access to this object data at the same time.

The direct opposite position is raw-data or also called early-fusion. In this approach minimally processed data from multiple and even diverse sensors are correlated in feature extraction levels to form a common environmental model or at least classification result. This method demands extensive data processing which directly converts into more complex hardware requirements. It reacts more sensitive to minor changes in the individual sensor and thus is less flexible compared to high level fusion methodologies with reference to changes of the sensor setup. The data interface between supplier and OEM is defined by the special raw data performance of the actual sensor and thus highly individual. Economically and strategically it appears as inferior to higher level fusion approaches. However, one argument may overrule the others in special cases: raw data fusion has potentially the highest detection and perception performance among all fusion concepts. Hence, economically its utilisation is justified, if the required environmental sensing performance can definitely not be achieved by other means. One prominent example for that is a smart night view approach described in section 7.

As a consequence, the best solution for future systems is a concept which will utilize less processed data than the object level concept but offers more strategic and economical opportunities like the early fusion concept does. Such a concept is mid-level fusion where fusion is performed on target level. This compromise would render possibility to convert the companys´ own knowledge and competences into an outperforming environmental representation and hence competitive advantage.

For most of the future safety and comfort functions one challenge will be the maximum exploitation of data from different sources of object information from diverse sensors with different quality of data to provide a maximum precise environmental representation. Another challenge is to identify a fusion approach which could be structured in a fusion module to operate as multi applicative basis to prevent redundant development efforts.

The target level or also called feature/mid-level fusion approach appears as the best compromise. The interface between supplier and OEM in principle remains as well proven for the object level today. The mid-level approach is close to the high level fusion, but offers innovative and differentiation potential. Hence, it offers the maximum unity in diversity in terms of information content, economical benefits and technical requirements. Sections 5-8 demonstrate some examples and emphasize essential issues for future fusion architectures.

In summary, there is no one solution for all, but target level fusion is one approach that provides a maximum unity in diversity for future requirements.

4. Modular sensor data fusion architecture

To achieve re-use of software modules and come to more standardized building blocks, the definition of sensor data fusion architecture including modules and interfaces is necessary. Figure 1 depicts a suitable approach for the sensor and the signal processing layer.

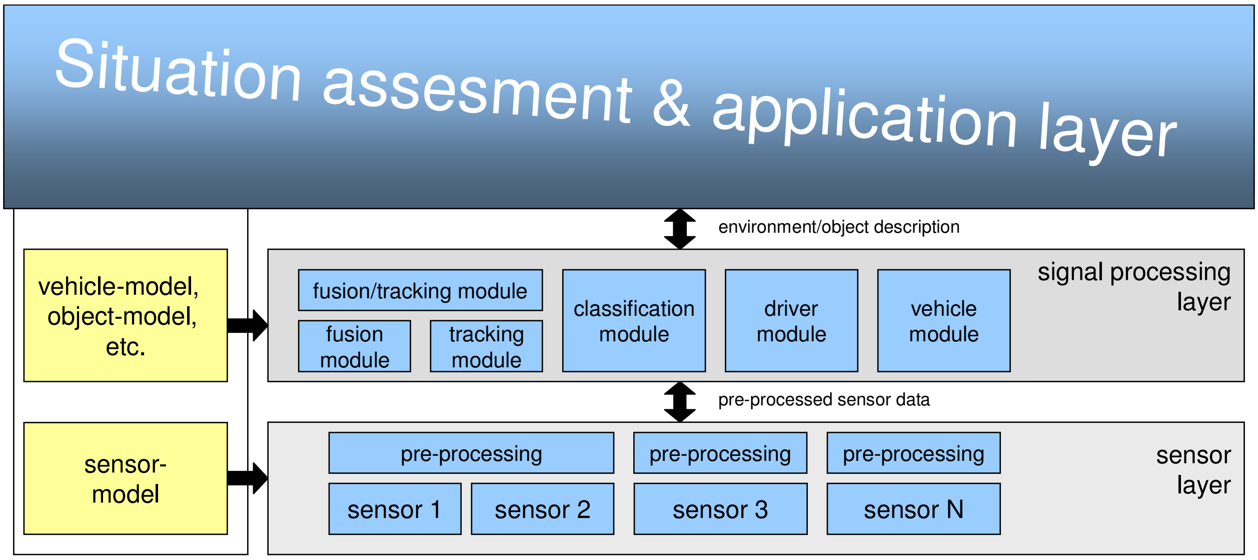

Figure 1: Overview of modular architecture

Pre-processing steps for the various possible sensor types like radar, lidar and camera sensor have to be developed and tailored for each individual sensor separately while keeping some of the basic processing steps like coordinate transformations, etc. as unified components. Practical experience from many projects has shown that the detailed understanding of sensor characteristics helps to achieve increased performance of the environmental description in terms of false and missed detection rates and accuracy of the object state estimation.

Key components of the signal processing layer are sensor data association, gating and filter methods which can be applied on target or object level of the sensor inputs from the sensor layer. Different levels of sensor data fusion like track to track fusion, mid-level fusion or even low-level fusion approaches can be implemented here taking into account, that different sub-sets of interfaces have to be defined dependent on the chosen fusion approach. Therefore also mapping or grid computation components belong to this processing layer.

An example for a concrete realization of a fusion framework is sketched in figure 2.

Heterogeneous sensor sources, in detail short range radar, long range radar and camera sensor, are fed into the fusion framework. After a sensor specific pre-processing of the targets/objects, the data has to be aligned in time and space. That is the data input representation for a Kalman-tracker which is able to handle point and extended objects at the same time.

To fulfil the requirements derived from different situation analysis / application constraints, the output of the Kalman-tracker stage has to be post-processed. Object selection, information accumulation and time prediction of object states are implemented in this processing layer and transferred to following steps, which try to interpret the delivered environmental description to come to situation awareness.

It has to be mentioned, that there are constraints concerning the use of a single sensor data fusion framework implementation for multiple applications. At first step, the concept aims at the integration of processing for a class of applications. Taking into account, that different classes of applications often leads to totally different or in worth case even contradictory requirements which can not be met with a single implementation. But still in this case the use of defined component and interface structures helps to speed up the development and leads to lower failure rates of the overall system.

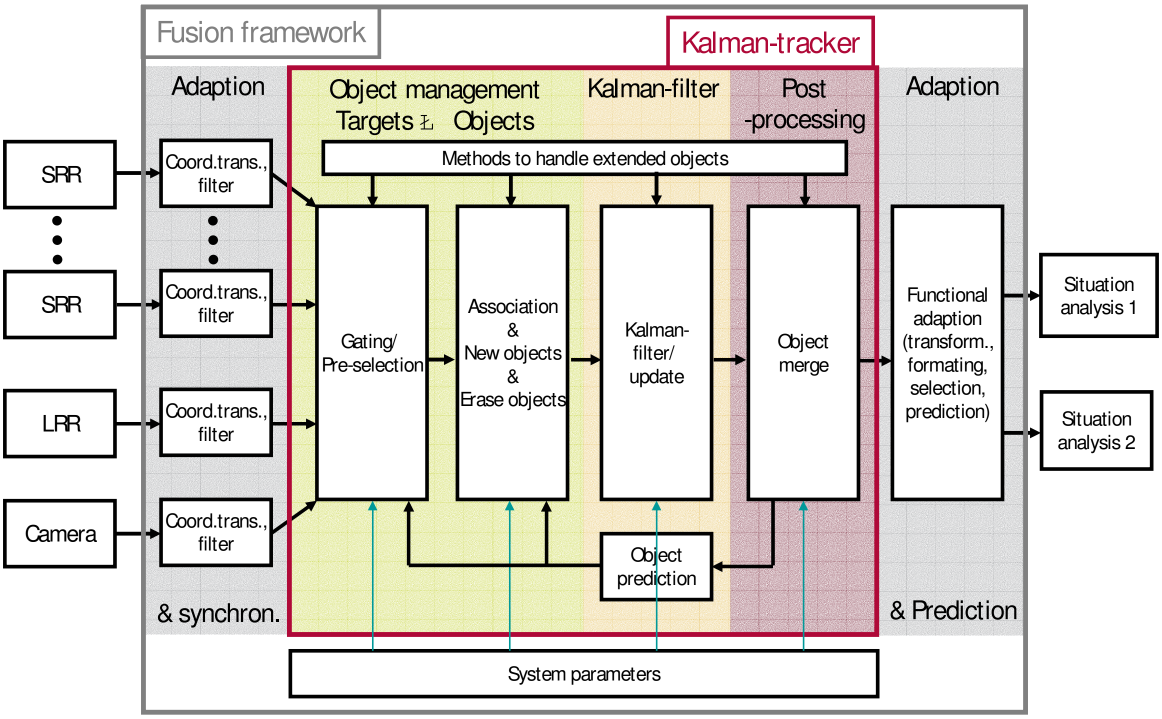

Figure 2: Fusion framework approach for multiple applications and modularized algorithm design

5. Fusion of Laserscanner and Short Range Radars

Scanning laser devices emit light rays in the non-visible region over their field of view. The distance to objects in the environment is could be calculated from the time of flight of the reflected rays. Because the angles between the light rays range from few degrees to a fractional amount of a degree, respectively, a multitude of reflections originates from a real world object, which leads to complex environmental scans showing the contours of real objects, see Figure 3.

This figure shows two scenes recorded in real traffic. Small dots represent single echos from the laserscanner, the bigger circles originate from radar targets. For illustration purpose, the environmental scans are captured along with images from a video camera looking at the scene. In contrast to a camera image, an environmental laser scan directly provide range information of objects. A big advantage of laser scanners over radar sensors is the ability to deliver extended targets instead of point targets. This advantage especially becomes apparent when looking at the right image in Figure 3, where large parts of the truck are simply not visible to the radar sensors, whereas the laser scanner perceives its dimensions.

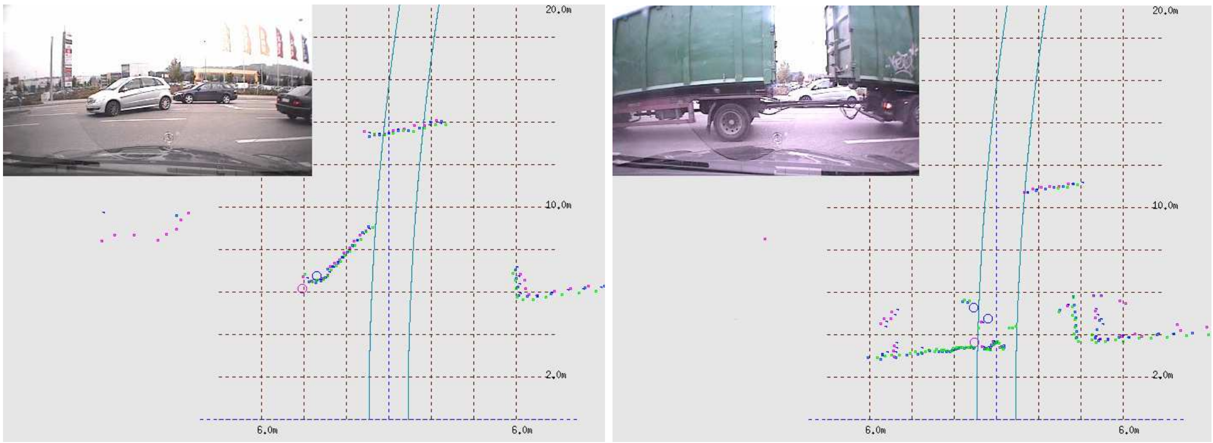

Figure 3: Environmental scan examples. Single laser echos are shown as points, the bigger circles represent radar targets.

Due to the optical measuring principle, the detection of ghost targets cannot be excluded in general. They may arise from bad weather conditions like heavy snowfall or heavy rain or from steam columns ascending from street drains, for example. The fusion with short range radars can suppress these kinds of ghost targets. In turn, short range radars could also generate false detections (e.g. flat metallic objects). In this case, the formation of a false object would be prevented by the laser scanner not detecting the object. An advantage of short range radars is that they can measure the velocity of objects directly by evaluating Doppler shift. Laser scanners provide information about object dimensions combined with high accuracy in position estimation, properties that are required for developing assistance functions for more complex environments like intersections or inner-city areas.

Our experimental setup consists of a laser scanner manufactured by the IBEO company (Alasca XT) and two 24 GHz short range radars. The laser scanner is characterized by its large field of view (160 degrees, range: 200m) and its range accuracy. The distance can be measured inch-perfect, the angular resolution is one degree at the outer of the field of view and decreases stepwise towards the central axis. Moreover, the measuring range of the scanner is horizontally enlarged to 3.2 degrees and divided into four scan planes, enabling the compensation of the vehicle pitch motion without losing the object. These properties enable the laser scanner to handle complex environments very precisely. The short range radars measure in a range up to 30m and have a field of view of 80 degrees each, where the angular accuracy decreases towards the boundary.

Measurement data of short range radars and laser scanner are combined within midlevel fusion which means a fusion on target level (=feature level). A categorization of different fusion concepts can be found in [4].

Figure 4 shows the architecture of the system. In the pre-processing step, raw scan data are subdivided into segments of adjacent points. These segments serve as laser scanner targets. Radar targets describe merged adjacent peaks, representing one reflecting object. A fusion on signal level is not feasible due to missing correlations in raw data because of the different sensor technologies. Single laser reflections cannot be fused with radar signals representing a point target. Fusing on target level systematically combines the advantages of each sensor, that is, on the one hand, the precise position estimation provided by the laser scanner, and on the other hand the velocity estimation provided by the radar sensors.

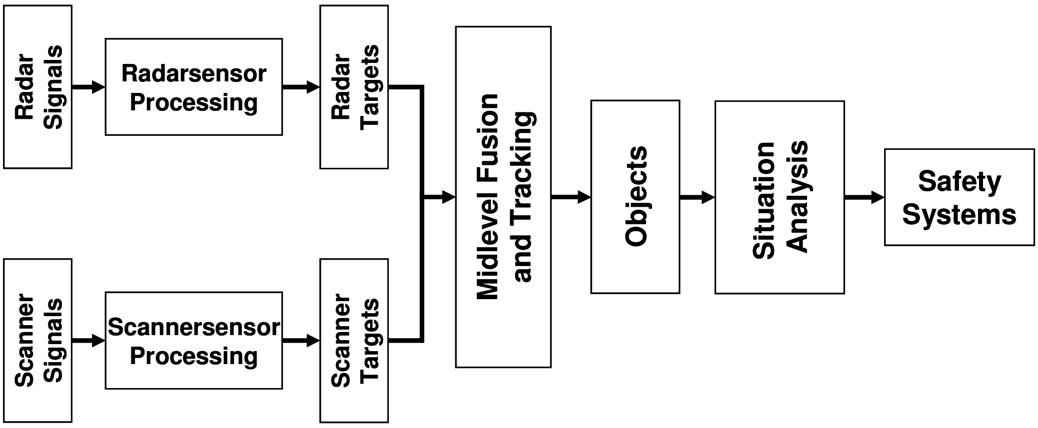

Figure 4: Target-Level fusion architecture

Figure 5 visualizes the necessary steps of object formation by means of an example. The real world object is detected by both types of sensors, laser scanner and short range radars. The pre processing step of laser raw data builds a segment out of the single echos belonging to the contour of the cube. The approach is gridbased and uses the CCL-algorithm (Connected Components Labeling) to merge occupied grid cells. A more detailed description is given in [4]. On target level, the fusion of the measurement vectors of laserscanner segments and radar targets takes place. Measurement vectors contain the position of the detected objects. In case of laser scanner segments, the position is given via a reference point, e. g. the center of gravity. For object tracking and object formation, well-known Kalman-Filter techniques are used [5].

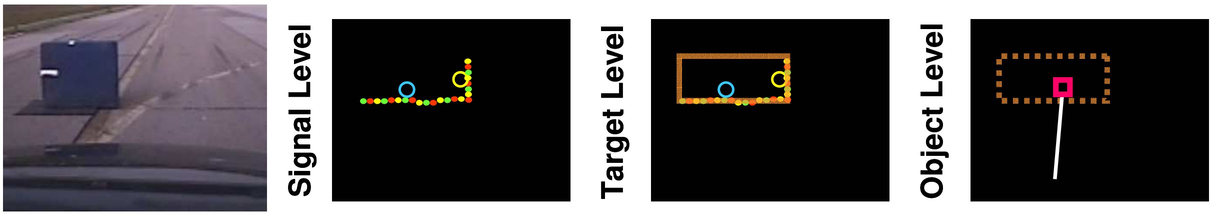

Figure 5: Object formation from Laser-Scanner and radar raw data

The property of the laser scanner to detect expanded objects necessitates to investigate in special cases in the fusion step. If radar targets are located within a segment box, but far away from the reference point, the measurement vectors of both sensors are not fused in order to avoid high jitter in object tracking. Nevertheless, the information about the object being detected by the radar sensor at all, is not discarded, but can be used for object validation, afterwards.

The fusion approach using laser scanner and short range radars was validated within a demonstration pre-crash application. Experiments comprise complex test scenarios conducted at special test areas as well as driving in normal traffic on highways, rural roads and in urban areas. Test drives in normal traffic were performed during day time to cover representative traffic situations and were partly performed under adverse weather conditions like rain, fog, wet roads and traffic spray. The system did not produce any false alarms covering a distance of 1600km in total. Furthermore, the system was evaluated in comprehensive crash and non-crash test scenes. A detailed description of missed alarm and false alarm rates and an overview of the different test scenarios can be found in [6].

Figure 6 points out the advantage of a large field of view as provided by the laser scanner.

Passing a cube at a lateral distance of two meters, the object stays in sight of the laser scanner almost until the vehicle's front passes by whereas the radar sensor loses track at a distance of over three meters.

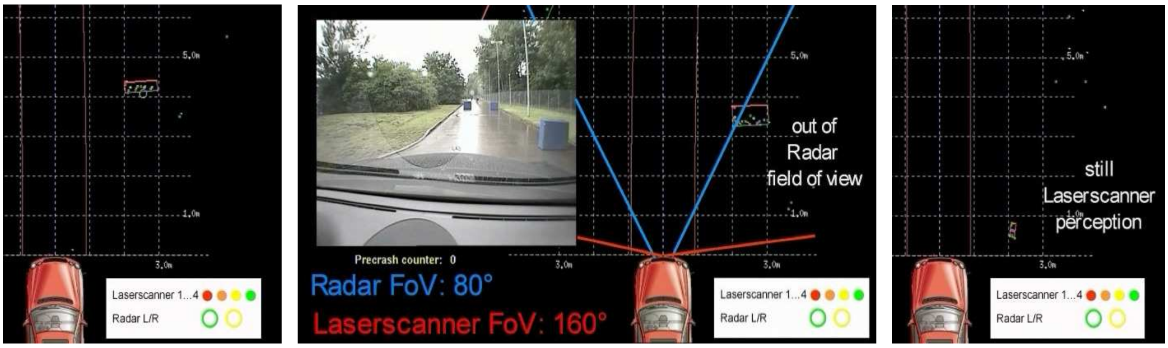

Figure 6: Comparison of different field of views.

In summary, target level fusion could be applied to divers sensors like laser scanner and radar and has demonstrated to support challenging functions like pre-crash.

6. Sensor Fusion of Lidar and Mono-Vision

Today's single sensor ACC systems in general suffer from the inability to detect stationary vehicles because the most discriminating detection feature is the velocity magnitude. Therefore systems on the market offer a "follow-to-stop-and-go" assistance. The different behaviour for stopping vehicles and invariably stationary vehicles is often incomprehensible to the customer and difficult to explain [7]. In order to deal with these aspects, a target-level fusion approach consisting of a serial monocular video camera and a serial lidar with a Field-of-View (FOV) of 16° is proposed. It combines precise distance measurements from the ranging sensor with the much better lateral object localisation capabilities and the strong texture-based detection features of the video sensor (Figure 7).

Fig. 7: Test vehicle with serial Night View Assist camera and multi-beam lidar sensor.

A requirement for each sensor fusion setup is the spatial and temporal sensor alignment. A hardware synchronisation of both sensors was chosen, as is the optimal choice with respect to inter-sensor data association and computation time aspects. However, all future sensors intended for sensor fusion systems must provide any mechanism for temporal alignment, whether it is an external measurement trigger signal input or the generation of timestamps relative to an external master clock in a free running mode. The spatial alignment is computed with a novel calibration procedure published in [8]. Using this alignment, the signal processing stage first projects the lidar measurements into the image domain. The image regions containing a lidar echo are further processed with a cascaded Boosting classifier based on Haar-like features [9]. The combined feature space of the lidar amplitudes and the image classifier response is used for the computation of existence probabilities per detection [8] (Figure 8).

The tracking stage processes these detections using latest tracking algorithms for simultaneous minimization of state and existence uncertainties [10]. In the state and existence prediction phase, it incorporates prior state constraints like sensor FOV and occlusion reasoning as well as a coordinated turn motion model for the sensor host vehicle and the observed vehicles. The sensor update step processes the polar lidar coordinates of a detection, its image positions and its sensory existence probabilities in a target-level fusion approach (Figure 8).

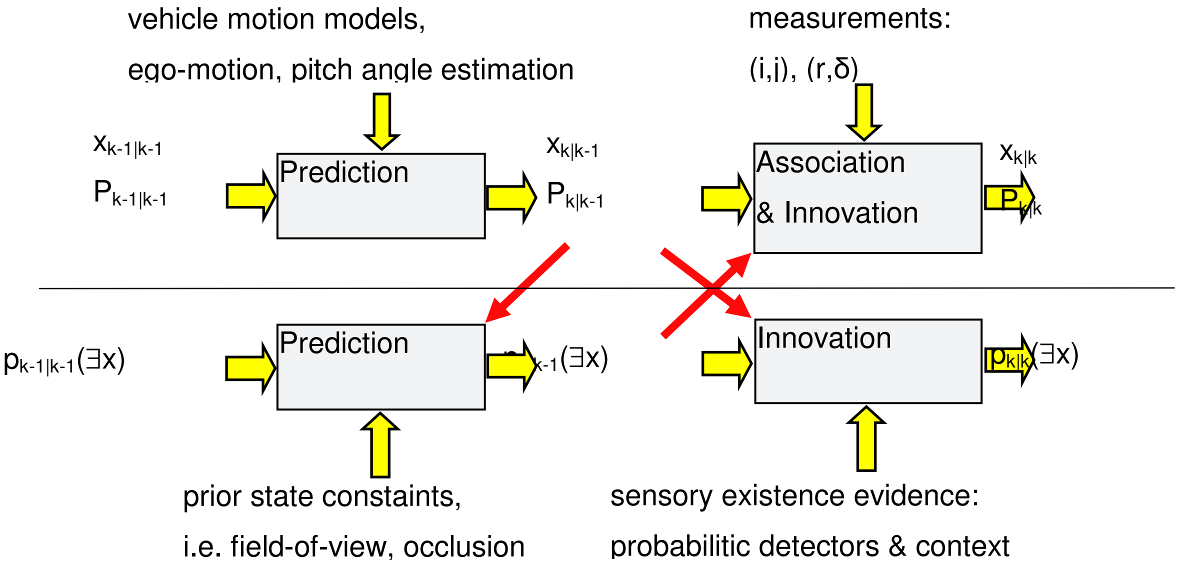

Fig. 8 – Feature level fusion approach for state & existence

Using this approach in both domains, existence and state, has the advantage of a combined and richer feature space for decision making. Furthermore, the environment model of a centralized tracking approach has tighter state enclosures, and less existence uncertainties compared to the de-centralized approach, due to the fact that it represents the past measurement information of all sensors and not a single one. This more accurate knowledge of the past leads to less ambiguous data association in each sensors measurement space. In addition to discriminating dynamical object behaviour features and in replacement for innovation counting heuristics, the explicit modelling of object existence probabilities enables the systematic and mathematically funded incorporation of measurable appearance features and the suppression of temporary detection failures for a Bayes-optimal time smoothed detection decision [10]. This is an extension of the approach in section 5 for a more robust decision.

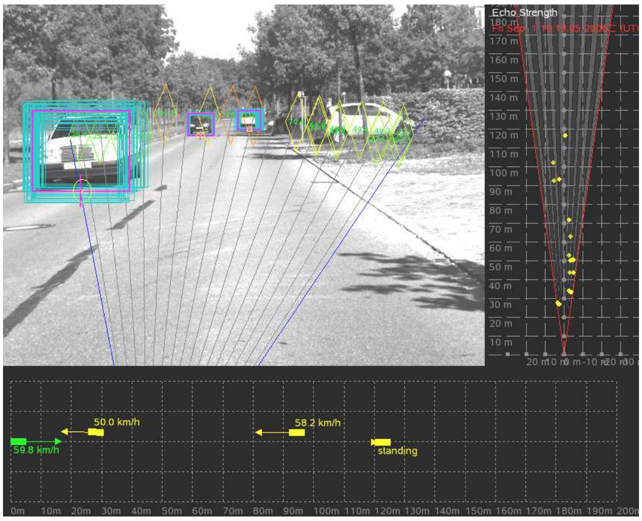

Fig. 9 – Signal processing, detection and tracking of ACC control targets.

The system is implemented in a demonstration vehicle with serial mono vision camera and a serial multi beam lidar sensor. Even stationary vehicles are detected in distances up to 150 meters and the system can keep track of already detected vehicles up to 200 meters. The current development phase allows non-synthetic in-vehicle demonstrations in real traffic scenarios (Figure 9).

7. Buffering or “Out-of-Sequence-Measurement”

For pre-crash applications a multiple sensor mid-level fusion system with a signal processing based on target level is favourable. Due to a sensor setup with components being manufactured by different suppliers, incoming measurements have a non-standardized interface with different time delays and varying time cycles, that are not synchronized to each other. The main challenge is to treat the out-of-sequence measurement (OOSM), resulting in the problem of how to update the current state estimates with measurements that arrive “too” late. In order to process OOSM, a sufficiently accurate reference time for all sensors is crucial. The integration of multiple heterogeneous sensors has increased the demand for handling asynchronous data fusion. Figure 10 illustrates an example of the OOSM problem: sensor 1 aquires the target earlier than sensor 2, but due to a larger latency on the interface, these data arrives at the processing unit even later than the targets acquired by sensor2. Simple Buffering (i.e. deferring the object fusion, till the data of the slowest path arrives) is in the most time critical applications such as pre-crash not applicable. This would mean, depending on the sensor mix, that fusion results could be delayed by several 10 ms. This is crucial in urban pre-crash situations.

An introduction to OOSM can be found in [11]. It was shown, that simple buffering can only be competitive to advanced algorithm with certain qualifications for the time update rate. In [12] different synchronization strategies are introduced and the worst case measurement latency is determined. In [13] Bar-Shalom presents an optimal algorithm for one-lag tracking in case of linear systems. An extension of the one-lag algorithms - the so called multi-lag - has been explored in [14] and [15]. Over the last few years, OOSM has intensively been investigated but there is still a wide area of open issues. Altogether, future research activities should focus on developing new algorithms in order to incorporate OOSM with simultaneous consideration of the scalability of the tracking system.

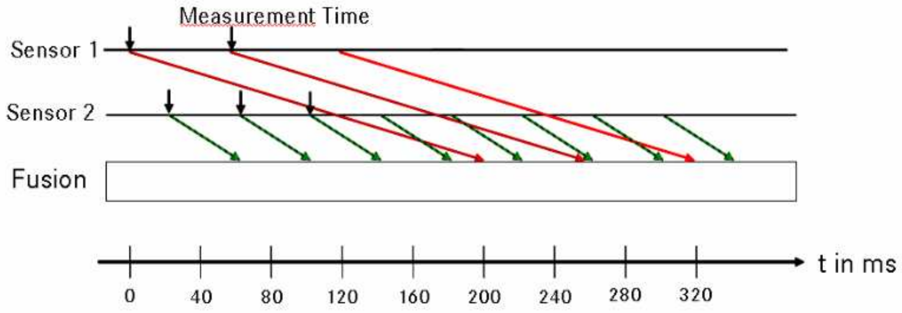

Figure 10: Out-of-Sequence Measurements arriving at the fusion center

The signal processing units consists of the following steps. The data acquisition describes the gathering of data from two radar sensors working on different clock rates, one can be polled the other is free running. Let us assume the time cycle for sensor 1 and sensor 2 is 66ms and 40ms, respectively. We do not investigate the use of one sensor to trigger the filter steps, we want to update out of sequence. Under this assumption, the situation in figure 10 can occur.

The question is how to include a measurement that was produced at an earlier time into the filter which already has been updated with measurements at later time steps. In the time prediction and filtering step, we are using a conventional Kalman Filter in the linear, extended and unscented form. Finally the data association is the most crucial part of a tracking framework. It is implemented with the auction algorithms as a global nearest neighbour and extracts pairs of new incoming and predicted measurements.

For us, the most promising approach seems to be the development of a combination of buffering and OOSM. In the future we will present the merits of incorporating measurements with different delays in a real world scenario with real world data. This will pave the way to deal with asynchronous data in future pre-crash systems.

8. Low-Level Fusion

Upcoming driver assistance systems not only require a high certainty of detection but also a high detection rate in order to get a complete perception of the vehicle environment. Within the project NIRWARN of the Ministry for Education and Research (BMBF) a pedestrian recognition system has been developed detecting pedestrians up to 90m on country roads. The pedestrians are highlighted in the camera image which is shown on a display mounted at the dashboard. To meet the desire to direct the driver's attention towards the pedestrians by a stronger warning a range of detection of 120m (equal to 4 seconds until collision at 100km/h) is necessary. In order to fulfill the demands of "faster, higher, stronger" in the high-end segment of driver assistance systems one single sensor is not sufficient any more. Also a fusion on object level is not sufficient: increasing the range of detection influences the certainty of detection negatively. Such a task is too complex for a single sensor system. The goal, however, is a high rate and a high certainty of detection at the same time. The feature level fusion combines object features offering the ability of getting more significant representations of the objects which are searched for. This approach is able to cope with the complexity of the detection task. To this end, each sensor extracts features which are processed by a detection stage in a common feature space. Thus, the information contained in the single physically different features is considered directly within the fusion process (MultiSensorBoosting). In particular, the combination of features yields a higher information content. In other words more significant feature structures arise which are not considered on an object level fusion.

First investigations in the context the follow-up project PROPEDES of the Ministry for Education and Research (BMBF) on a fusion of a far infrared and a near infrared camera for the detection of pedestrians at night confirm this fact. As can be seen in table 1 the fusion detector needs less features to differentiate between a pedestrian and the background than each other single detector.

	Number of features
Detection only Near infrared (NIR)	**734**
Detection only Far infrared (FIR)	**426**
Fusion result (NIR & FIR)	**196**

Table 1: Number of necessary features for the detection of pedestrians. The fusion detector needs less features to differentiate between a pedestrian and the background than each other single detector.

The individual features as well as their combinations were not selected and parameterised by hand but statistically based on a learn set using the AdaBoost algorithm. Figure 11 shows the selected features of an individual detection stage: in the FIR image large-scale features are selected reflecting the thermal contrast between the pedestrian and the background whereas in the NIR image relatively small features are selected representing the structures (arms and legs) of a pedestrian. The single features are insufficient for a reliable pedestrian recognition. Only the combination of both features in the fusion step fulfills the high demands on the detection and false-alarm rate.

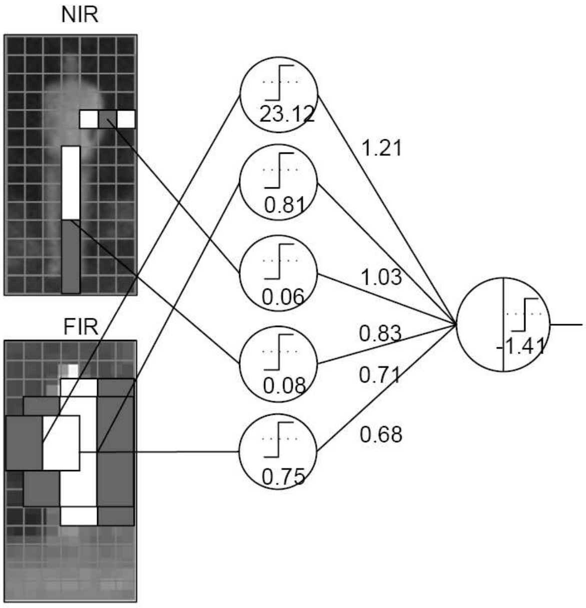

Figure 11: Exemplary features of a detection stage. The combination of the features increases the significance compared to the single features.

The results of the fusion detector using a test set show that the fusion detector outperforms the single detectors (see fig. 12). Due to the fusion on feature level a better detection performance can be reached with less features reducing the computational burden. Thus the overall costs of such a driver assistance system stay economically in spite of additional sensors.

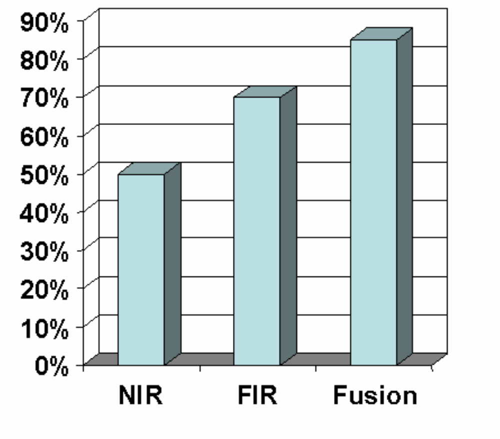

Figure 12: Detection rates with fixed false-alarm rate. The fusion detector reaches - in spite of less used features - a higher detection rate than each other single detector.

9. Summary

With the intensive use of optical sensors for safety and comfort functions prompts the question for the optimum fusion approach to exploit the huge information content available in optical sensors. There is no one for all fusion solution. There could be future applications which might require low level fusion, especially if video based classification is the main purpose. On the other hand, mid-level fusion appears to have the potential as a general approach to meet cost, strategic and performance constraints for future automotive applications.

10. Publications

[1] Schneider R., Wenger J., "Advanced Microsystems for Automotive Applications 2003", in VDI-Buch, Springer Berlin Heidelberg, ISBN-978-3-540-00597-1, 2003, pp. 97-111

[2] De la Torre, F.; Vinyals, Gavrila D, "Learning Kernel Expansions for Image Classification", Computer Vision and Pattern Recognition, 2007. CVPR apos;07. IEEE Conference on Volume , Issue , 17-22 June 2007, pp.1 - 7

[3] Stein F., "Efficient Computation of Optical Flow Using the Census Transform", Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2004, ISBN-978-3-540-22945-2, pp.79-86

[4] Skutek, M.: Ein PreCrash-System auf Basis multisensorieller Umgebungserfassung. TU Chemnitz Diss. 2006

[5] Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems, in Transaction of the ASME – Journal of Basic Engineering, 1960, vol. 82, pp. 35-45

[6] Pietzsch, S., Aycard, O., Burlet, J., Vu, T.D., Hackbarth, T., Appenrodt, N., Dickmann, J. and Radig, B.: Results of a Precrash Application based on Laserscanner and Short Range Radars, accepted at IEEE Intelligent Vehicles Symposium, June 2008, Eindhoven, NL -to be published-

[7] A. Neukum, T. Lübbecke, H.-P. Krüger, C. Mayser, J. Steinle: „ACC-Stop&Go: Fahrverhalten an funktionalen Systemgrenzen", In 5. Workshop Fahrerassistenzsysteme (FAS), Walting im Altmühltal, 2008

[8] M.Maehlisch, R.Schweiger, W.Ritter, and K.C.J.Dietmayer, "Sensorfusion using spatio-temporal aligned video and lidar for improved vehicle detection," in Proceedings of IEEE Intelligent Vehicles Symposium, Tokyo, Japan, 2006.

[9] P.Viola and M.Jones, "Robust real-time object detection," in Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, Vancouver, Canada, July 13, 2001.

[10] Mirko Mählisch, Magdalena Szczot, Otto Löhlein, Michael Munz, Klaus Dietmayer: "Simultaneous Processing of Multitarget State Measurements and Object Individual Sensory Existence Evidence with the Joint Integrated Probabilistic Data Association Filter", 5th International Workshop on Intelligent Transportation, Hamburg, 2008

[11] Out-of-Sequence Measurment Treatment in Sensor Fusion Applications: Buffering versus Advance Algorithms. Moritz Mauthner, Wilfried Elmenreich, Alexander Kirchner and Diego Boesel. Workshop Fahrerassistenzsysteme 2006

[12] Data synchronization strategies for multi-sensor fusion. Nico Kaempchen and Klaus Dietmayer. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, 2003

[13] Update with out-of-sequence measurement in tracking: Exact solution. Y.Bar-Shalom. IEEE Transactions on Aerospace and Electronic Systems, 2002.

[14] Advances in asynchronous and decentralized estimation. M.Mallick, S.Coraluppi and C.Carthel. IEEE Aerospace Conference 2001.

[15] Multi-sensor multi-target tracking using out-of-sequence measurements. M.Mallick, J.Krant and Y.Bar Shalom. Fifth International Conference on Information Fusion 2002.

Workshop on Sensor and Data Fusion – Architecture and Algorithms
Nov. 20., 2009, Berlin, Germany (invited contribution)

Multi-Level Fusion for Cooperative Cognitive Automobiles

Christoph Stiller
Institut für Mess- und Regelungstechnik, Karlsruhe Institute of Technology (KIT), Germany

***Abstract*— This contribution reports on the fusion levels that are employed within the Karlsruhe-Munich Transregional Collaborative Research Centre *Cognitive Automobiles.* Automotive applications impose challenging requirements on the reliability of environment information as achieved by automotive video, radar, or lidar sensors. This contribution approaches perceptual reliability from the viewpoint of fusion of diverse information at various levels.**

I. INTRODUCTION

Among the most fascinating capabilities of intelligent beings is the seamless perception and interaction with their environment. Guidance and control of automobiles comprises a comprehensive example for these capabilities. Based on the understanding of the scene human drivers plan, initiate, supervise, and control suitable behavior. Driver assistance systems aim to project those capabilities onto artificial systems. At the very extreme end of driver assistance ranges the complete automatic control of automobiles. Beyond encouraging successes in the research of automatic driving, unsupervised automatic driving in public traffic is still a far fetched vision. One of the major limiting shortcomings of driver assistance systems is their lack to reliably identify those situations in that sufficient performance cannot be guaranteed.

This contribution concerns itself with techniques to enhance the perceptual reliability of automotive environment perception beyond its current state. These techniques can be considered being based on an identical principle, namely the fusion of information through the exploitation of diversity at various levels.

The work is embedded in the Karlsruhe-Munich Collaborative Transregional Research Centre Cognitive Automobiles of the DFG comprising the following partners:

- Karlsruhe Institute of Technology (KIT),
- Fraunhofer Institut IITB Karlsruhe,
- Technische Universität München,
- Universität der Bundeswehr München.

It focuses on systematic and interdisciplinary research on machine cognition of mobile systems as the basis for a scientific theory of automated machine behavior. The potential of cooperative perception and behavior is examined. Analytic research is accompanied by closed-loop simulations. Experimental autonomous vehicles build an important platform for the TCRC that allows demonstration and validation of the theoretical findings.

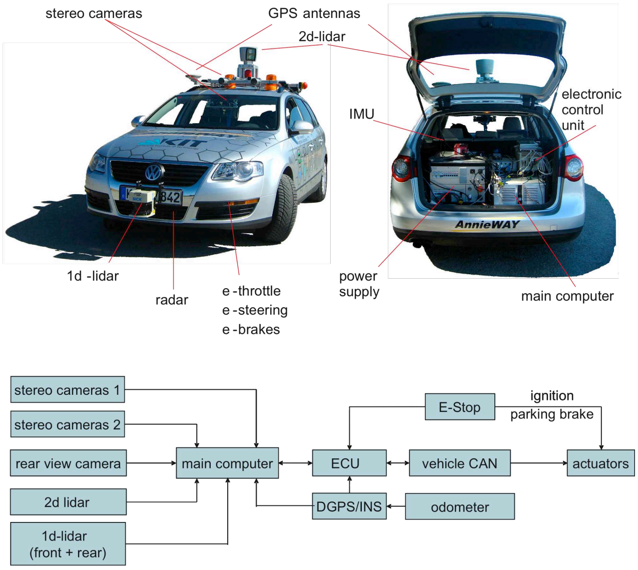

Fig. 1: Experimental vehicle set up

The same partners also contribute to the team AnnieWAY in the Urban Challenge 2007 competition [7]. Figure 1 depicts the vehicle

AnnieWAY that is among the six experimental vehicles of the project and hs reached the finals of the Urban Challenge.

II. DIVERSITY LEVELS OF COOPERATIVE COGNITIVE AUTOMOBILES

Cooperative Cognitive Automobiles exhibit a wide range of diversity in the information available.

A. Multiple Sequential Measurements

Typically, sensor sampling rates in the automotive domain are significantly higher than the Nyquist rate. In particular, much of the information in the environment of an automobile projects the static environment. Hence, huge redundancy is included in the temporal direction of the sensor signals that may be exploited through appropriate spatio-temporal registration.

B. Multiple Visual Keys

Numerous features have been exploited for visual perception of objects in the environment of automobiles in literature. These may be separated into the following two main classes.

The first class comprises features that may be extracted from a single image of a single camera. Such features will be referred to as *appearance features* and include color, edge activity, shape, shadow, dark wheels, and symmetry (cf., e.g., [3, 4, 5, 6, 14, 15]). The selection of specific features or a combination thereof is commonly chosen ad hoc aiming at identification of patterns and features that are 'typical' for the expected kind of obstacles. 3D information that is often needed for driving support functions may be acquired through feature tracking over time [1]. However, in unfavorable conditions, the individual features will be degraded or even not visible at all. While, e.g. the dark region under vehicles, i.e. the shadow, is a prominent and powerful key for vehicle detection in most situations, this feature becomes unreliable in some situations such as sun rise, sun dawn, and night. Furthermore, methods for obstacle detection based on appearance features are restricted to a predefined set of 'typical' obstacles. General appearance features for arbitrary obstacles cannot be imposed. This results in an insufficient reliability of such approaches inhibiting realization of some important safety functions.

The second class of features is based on correspondences between multiple images, i.e. sets of positions in the image that represent projections of identical points from the 3D real world onto the image plane. Such *correspondence features* include disparity of a stereo camera as well as 2D displacement in the image plane induced by motion of an object with respect to a single camera.

Correspondence features allow observation of 3D geometry and 3D motion of a scene and may therefore be considered more general features for obstacle detection as compared with appearance features. However, correspondence features are known to be restricted to obstacle detection at moderate distances or favorable object motion.

Obstacle detection techniques that exploit diversity by combining multiple of the above features may overcome the before mentioned restrictions. However, beyond promising examples for perception with multiple features [2, 4, 5, 6, 15], a general framework for a consistent combination of features is still an open research issue.

C. Multiple Levels of Perception

A totally different source of diversity stems from the process of perception itself. Figure 2 depicts a typical chain of the perception process.

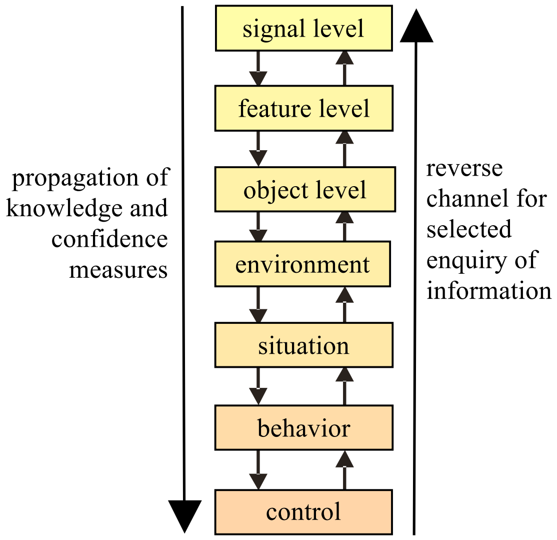

Fig. 2 Perception process

It emerges from the signal level (pixels) via feature and object detection, to environment and situation description and eventually reaches behavior generation and vehicle control.

While the main information flow is directed from the top towards the bottom of the chain, diversity can be augmented by additional usage of the reverse channel. Specific situations, e.g., may imply the presence of specific objects and its associated features or at least the overly likelihood of their presence. Thus information that is diverse to the image signal is propagated from the bottom towards the top of the perception chain. A probabilistic logic learning approach applying Markov logic networks has recently been proposed in [12]. It allows the inclusion of formal logic rules on the environment or situation level into the perception process. A complete and consistent usage of the reverse channel is still a challenging field of research.

D. *Multiple Sensors*

A further and obvious source of diversity is given by the information of diverse sensors. Figure 3 depicts a multisensor system that has been used for autonomous and unsupervised driving on assault courses [10].

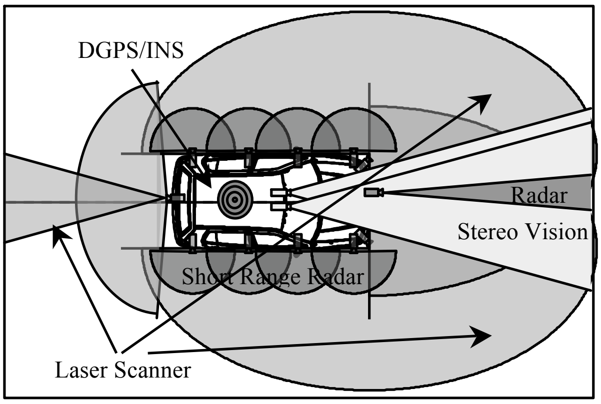

Fig. 3. Multisensor perception

Information fusion techniques (see e.g. [8] for an overview) not only yield a complete surround view but also allow for a safe function in case of failure or missing detection of any single sensor [10].

E. *Cooperative Perception*

As the equipment rate of vehicles with capabilities for environmental sensing increases, it becomes likely that a vehicle within a group possesses information about the environment that is relevant to others. This is illustrated at a specific scenario of a street crossing in Figure 4. The environment information perceived by the upper vehicle may not only enhance the reliability of object perception for the leftmost vehicle in the overlapping area but is indispensable for its object detection in areas that are outside the field of perception due to the limited field of view or due to occlusions. In those areas cooperative perception of multiple vehicles that communicate their information provides another attractive grade of diversity. It is worth noting that cooperative perception does not require a 100% equipment rate, but provides benefit even at moderate rates.

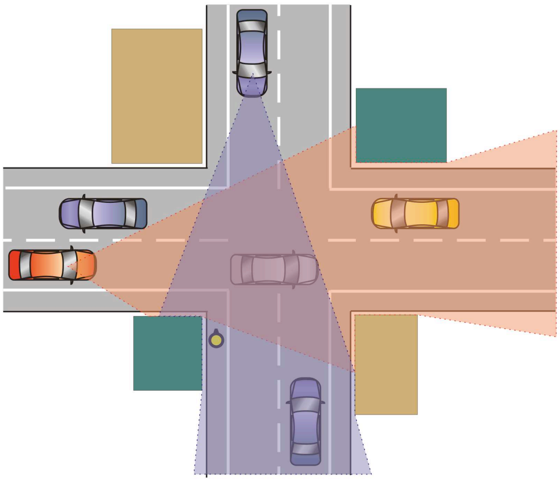

Fig. 4. A scenario for cooperative environmental perception in mixed traffic

Preliminary experiments with cooperative perception between vehicles have been reported in [13]. An important issue in this context is the spatio-temporal registration of data transmitted in the coordinate system of other vehicles. Since the uncertainty of the spatio-temporal alignment adds itself to the intrinsic uncertainty of the sensor information, this alignment must be conducted with high precision. It is shown that an alignment strategy that combines the coarse localization information of a GPS system with the sensor output of the video sensor itself yields good results for the envisaged application.

It is straightforward to expand this cooperation to the behavioral level, thus yielding a network of agents with a distributed decision structure.

III. SUMMARY AND CONCLUSIONS

This contribution has addressed the issue of perceptual reliability from the viewpoint of diversity. The rich information included in the vehicular sensor data raises the opportunity of fusion at various levels of diversity. While the cooperative exploitation of temporal diversity and diverse features as well as of information from diverse sensors has frequently been addressed in literature, little work has been reported so far on cooperative exploitation of information from diverse levels of the perception chain. As another field of research that is still in its early beginnings, cooperative sensing between vehicles in mixed traffic is outlined. Through utilization of diversity across multiple levels, environment perception may be conducted at a level of reliability that allows realization of safety functions.

ACKNOWLEDGEMENTS

The author gratefully acknowledges support of this work by the Deutsche Forschungsgemeinschaft (German Research Foundation) and the fruitful collobaration with the partners in Munich and Karlsruhe within the Transregional Collaborative Research Centre 28 'Cognitive Automobiles'.

REFERENCES

[1] Y. Bar-Shalom, “Multitarget-Multisensor-Tracking: Applications and Advances,” Boston Artech House, 2000.

[2] T. Dang, C. Hoffmann, and C. Stiller: “Fusing optical flow and stereo disparity for object tracking”. In Proc. IEEE Int. Conf. Intelligent Transportation Systems, Singapore, pp. 112-117, Sep. 2002.

[3] E.D. Dickmanns and T. Christians, “Relative 3-D state estimation for autonomous visual guidance of road vehicles,” in: 2. Conf. Intelligent Autonomous Systems, 1989.

[4] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Pätzold, and C. Wöhler, “Autonomous driving goes downtown,” IEEE Intelligent Systems, vol. 13, no. 6, pp. 40 - 48, Oct. 1998.

[5] M. Haag and H.-H. Nagel, “Combination of edge element and optical flow estimates for 3D-model-based vehicle tracking in traffic image sequences,” International Journal of Computer Vision, vol. 35, no. 3, pp. 295 - 319, 1999.

[6] C. Hoffmann, T. Dang, and C. Stiller: “Vehicle detection fusing 2D visual features“. In Proc. IEEE Intelligent Vehicles Symposium, Parma, Italy, pp. 280 – 285, June 2004.

[7] S. Kammel, J. Ziegler, B. Pitzer, M. Werling, T. Gindele, D. Jagzent, J. Schröder, M. Thuy, M. Goebl, F. von Hundelshausen, O. Pink, C. Frese, C. Stiller, „Team AnnieWAY’s autonomous system for the 2007 DARPA Urban Challenge.“ In: Journal of Field Robotics 25 (2008), September, No. 9, pp. 615 – 639.

[8] L. A. Klein, “Sensor and Data Fusion,” SPIE, Bellingham, Washington, 2004.

[9] H.-H. Nagel, “Steps toward a cognitive vision system,” AI-Magazine, vol. 25, no. 2, pp. 31-50, 2004.

[10] Ü. Özgüner, C. Stiller, and K. Redmill, Systems for Safety and Autonomous Behavior in Cars: The DARPA Grand Challenge Experience. In: IEEE Proceedings 95 (2007), Feb., No. 2, pp. 397–412.

[11] C. Stiller, J. Hipp, C. Rössig, and A. Ewald, “Multisensor obstacle detection and tracking“. Image and Vision Computing Journal, 18(5), pp. 389 – 396, Apr. 2000.

[12] C. Stiller, S. Kammel, I. Lulcheva, J. Ziegler, „Probabilistische Methoden in der Umfeldwahrnehmung Kognitiver Automobile. In: at – Automatisierungs-technik 56 (2008), November, No. 11, pp. 563 – 574.

[13] K. Tischler, M. Clauss, Y. Günter, N. Kämpchen, R. Schreier, and M. Stiegeler, “Networked Environment Description for Advanced Driver Assistance Systems,” IEEE Intelligent Transportation Systems Conference, Vienna, Sep. 2005.

[14] F. Thomanek, Visuelle Erkennung und Zustandsschätzung von mehreren Straßenfahrzeugen zur autonomen Fahrzeugführung, PhD thesis, University BW Munich, VDI Verlag, Fortschr.-Ber., ser. 12, no. 272, 1996.

[15] T. Zielke, M. Brauckmann, and W. von Seelen, Intensity and edge-based symmetry detection applied to car-following, in: ECCV, Lecture Notes in Computer Science, vol. 588, 865-873, 1992.

Data fusion for optical navigation systems

Anko Börner
German Aerospace Center (DLR), Institute for Robotics and Mechatronics,
anko.boerner@dlr.de

Abstract: Navigation systems mostly rely on multi sensor approaches. For example, the combination of GNSS and inertial measurement units is well understood and established. The inclusion of optical sensors in such systems opens up different possibilities to provide information being useful for navigation. Ego motion can be estimated even without or with incorrect absolute GNSS signals, 3D models of the environment can be determined by applying stereo sensors, context information out of monocular camera systems can provide auxiliary information using relevant object features.

The paper will introduce a number of systems and approaches with regard to optical navigation. The main part will focus on data fusion on different levels. Starting with synchronization and relative geometrical referencing, topics like calibration and complex filtering are covered. Promising results with regard to optical navigation systems will be presented. One of the major conclusions of our work in the last years is, that data fusion requires very precise knowledge about the overall system and all its components. The integration of several sensors with their dependencies with respect to different influencing factors creates a n-dimensional state space. Reliable state estimation fails, if system modelling and calibration do not meet the needed requirements.

Multi-camera systems, distributed systems and sensor networks: Architectures and algorithms for people surveillance

Rita. Cucchiara, Andrea Prati
Università di Modena e Reggio Emilia, Italy

ABSTRACT

After more than ten years of researches in computer vision-based video surveillance, results and solutions in single-camera systems are now extended into more complex architectures, with multi-camera platforms to provide multiple views of the scene, or with distributed systems for wide area surveillance often enriched by the capabilities of sensor networks. New architectures call for new algorithms as well, for addressing challenging problems such as consistent labeling or re-identification. This paper presents an overview of the design aspects of system architecture and computer-based algorithms for multi-camera, distributed camera and heterogeneous sensor-based surveillance systems, with specific regards to the problem of people surveillance. Models and algorithms for object segmentation and tracking in multi-camera and distributed systems will be presented, with reference to some projects of Modena's ImageLab.

1. INTRODUCTION

The explosion of requests of intelligent surveillance systems for different scenarios such as people monitoring in public areas, smart homes, urban traffic control, mobile application, and identity assessment for security and safety, leads research activity to explore many different dimensions in terms of both architectural issues and algorithms for scene recognition and interpretation.

Thus, many synergic fields spanning from hardware embedded systems, sensor networks, computer architecture on one side, and image processing, computer vision, pattern recognition on the other side, are synergically integrated to cope with real-time surveillance applications. Many sensor fusion and data fusion paradigms must be taken into account .

Surveillance *concerns models, techniques and systems for acquiring information about the 3D external world, detecting targets along the time and the space, recognizing interesting or dangerous situations, generating real-time alarms recording meaningful data about the controlled scene.* The nature and the richness of the information depend on the type of sensors: generally cameras are used since the informative content of image is normally higher than the one of other sensory data. Often audio is added in some specific contexts; for instance in monitoring soccer games the type of audio information are very discriminative for automatic event detection. Recently the possibility to collect in large network other less powerful but highly distributed sensors such as RFID, PIR or temperature sensors increased the range of sensory data which can be included in surveillance systems.

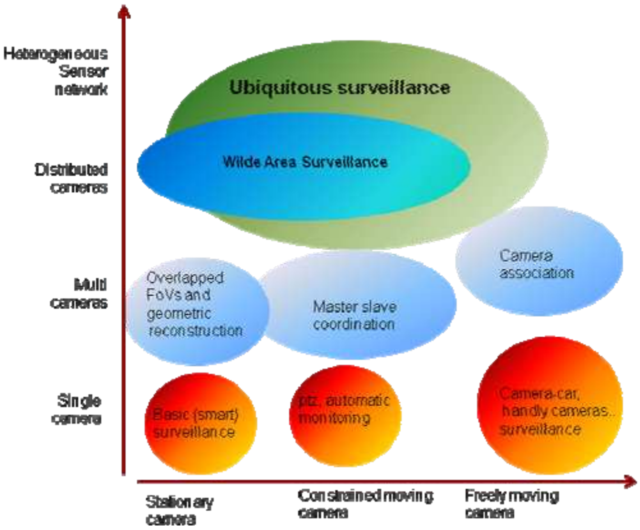

Fig. 1 Architecture of surveillance systems w.r.t. to acquisition modules

Design choices depend on the type of sensors; within standard visible light cameras, basic systems process a single data stream acquired by either a stationary (fixed) camera or a camera with a constrained motion such as PTZ camera or a camera with an unconstrained motion such as the ones mounted on vehicles or on wearable equipments (Fig. 1). Then, when multiple (heterogeneous) cameras are connected in a forest of sensors, standard techniques used in single- fixed camera surveillance are not sufficient anymore. Different approaches should be taken into account depending on the multiplicity of cameras, (e.g., with overlapped or not overlapped fields of view), the network capability and the client-server architecture. Other needs come for the request of merging heterogeneous sensors, as for instance RFIDs and cameras to identify and localize targets.

Future systems of ubiquitous surveillance will explore new solution in hardware and software to work with multiple data streams with heterogeneous cameras.

2. ARCHITECTURE AND ALGORITMS IN SINGLE CAMERA SURVEILLANCE SYSTEMS

The problem of discussing architectural aspects in video surveillance was often neglected, at least until some years ago, since the choices in hardware and architecture were only driven by the need of displaying good images in remote control centers or providing very simple and affordable image processing tasks.

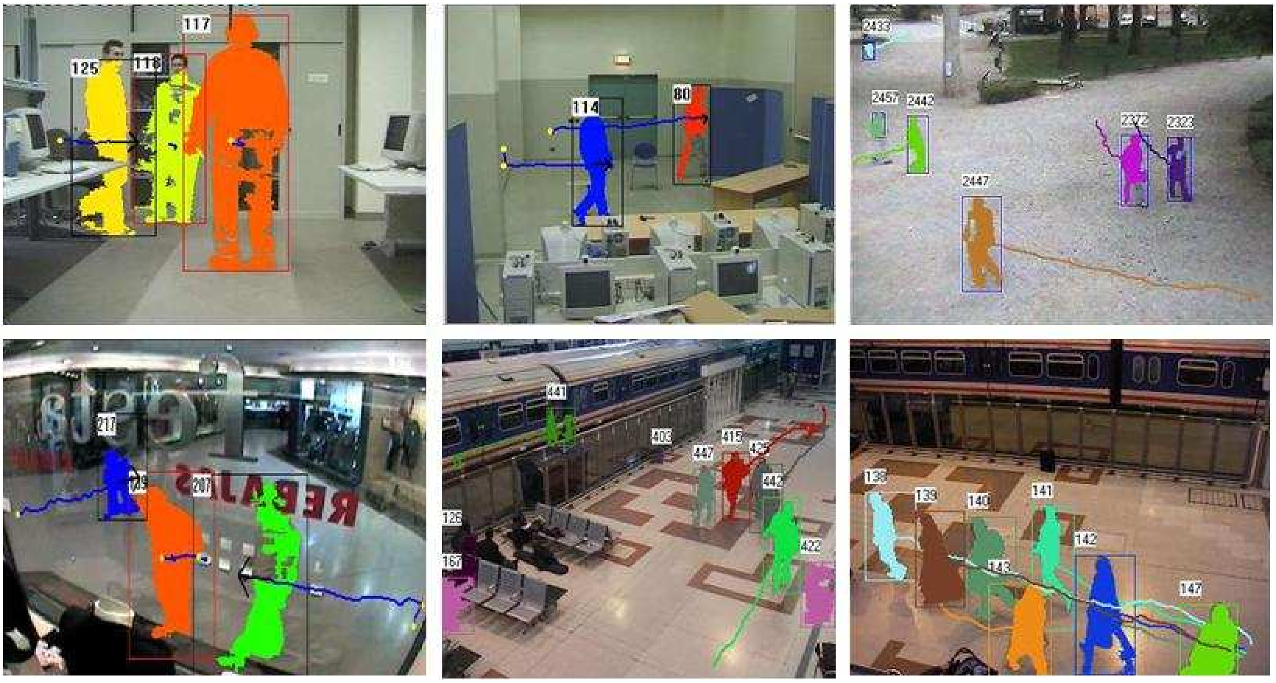

Fig. 2 Examples with Sakbot [6] and Ad-Hoc tracking in video from VISOR Repository www.openvisor.org (first row) and PETS dataset (second row)

Nowadays, the capability of low-level PCs and smart cameras, the bandwidth possibilities in network cameras and the availability of hardware and middleware for sensor network enlarged the architecture panorama.

Typical video surveillance systems consist at least of a single stationary camera, connected with a embedded or general purpose computer with local or remote storage, display resources and related computer vision software.

To increase the speed of surveillance task on standard computer the use of GP-GPUs is growing: for instance in [1] the author demonstrates that typical motion detectors on GPU may provide a speedup of 5 or more with respect to standard CPU of the same generation offering a comparison was between a P4-2GHZ and a GeForceFX 5800 Ultra (GF5800U) at 500 Mhz.

Few years ago the power of smart cameras or local systems was not enough to support the whole surveillance task so that the focus was in the compression and communication capabilities to transmit images also at low frame rate and low quality, but sufficient to be processed remotely. In [2] a model of performance analysis for video compression in surveillance was presented, showing that with an efficient H264-AVC code also using low bandwidth channel images was sufficiently detailed to provide segmentation and tracking of people and vehicles remotely. Conversely, new smart cameras can be exploited often to provide similar jobs at client side in distributed environments as it will be discussed further.

As it is well known, in surveillance some tasks are unavoidable, although their execution order and their mutual feedback can be designed in different ways, according with the architecture of the system and the peculiarities of the application. For instance some systems identify moving objects and then classify them to select only useful targets (e.g., vehicles and people); others, instead, provide target detection at the beginning and then track selected objects only. A part from some initial image enhancement and image processing steps, in any surveillance system we should have four steps whose order is not a priori defined.

1. Detection: *the task of exploiting space coherency to extract area of interest typically at image segmentation level.* In stationary camera it is always provided by background suppression method, while in more general scenarios with moving cameras or without a reference image, segmentation according with motion fields, color texture and so on have been adopted.

Detection with background suppression was initially well studied in traffic analysis, dating back to the pioneering work of 15 years ago of Koller et al. [3]; a good survey in this area is the work of Kastrinaki et al. [4]. Then it has been exploited for people and other objects in indoor and outdoor applications with the analysis of shadows and artifacts, initially in university-based projects such as W4 [5], Sakbot [6] and Knigth [7].

The most known model for background is the one based of mixture if Gaussians or MoG [8] used to cope with multimodality in the background model, but many others methods have been further proposed such as the one based on multilayer segmentation [9] or with kernel density estimator [10]. All these methods obtain region segmentation according with motion followed by some tasks of labeling, data selection and feature extraction.

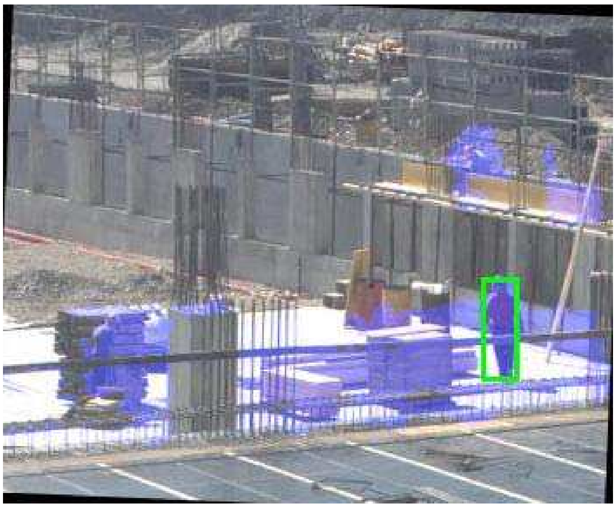

Fig. 3 An example of people detection [22]

2. Tracking: *the task of exploiting time coherency to follow the same object along the time and the space.* Here the dichotomy of tracking-by-detection or detection-by-tracking is still open [11]. When detection is easily provided in stationary camera, appearance-based tracking working at pixel level is preferable so that also the object shape is extracted. Example are [12], and the Ad-Hoc tracking we have defined and extended in multi-camera system [13], whose results are reported in Fig. 2.

However, in more general cases with PTZ or hand-held cameras other probabilistic methods must be used and in particular mean shift [14] or its variants (such as Camshift implemented in OpenCV) and particle filtering [15]. Complete surveys are [16, 17, 18] and an interesting work in performance evaluation is [19].

3. Recognition: *the task of exploiting model coherency to provide object identification and classification.* This problem was initially underestimated in surveillance since, once fixed the scenario, the type of objects was often pre-defined, such as vehicles in road or people in an office. Of course many pattern recognition systems have been exploited to classify these objects [20] after segmentation and tracking. Instead, there is an increase of research effort in model-based recognition for surveillance, especially for people recognition in cluttered environment. They exploit the knowledge of object models also in 3D to identify and localize them in videos or machine learning on visual features.[21]

An example of research in this direction is the recognition of people in construction working sites, where the presence of many moving regions, moving cameras and distractors make the basic background suppression methods infeasible. In this field, indeed, approaches of direct people detection in single images could be more affordable. An example is shown in Fig. 3 where a machine learning approach based on covariance matrices and a boosted classifier have been adopted [22].

4. Understanding: *the task of exploiting the world model to recognize actions, events and behaviors.* This is the final step, dependent on the surveillance application. New generation of smart surveillance systems must integrate a final step of high level reasoning to assess the situation, and infer possible dangerous or interesting behaviors, action or events. These tasks are similar in single or distributed systems and normally are provided as a centralized step. Typical examples are the analysis of abnormal behaviors based on trajectory clustering [23, 24]; if no information on the context are embedded, typically abnormal means infrequent and reasoning is based on the occurrence of the specific evidence.

3. MULTI-CAMERA AND DISTRIBUTED CAMERA SYSTEMS

The natural evolution of standard surveillance systems goes in the direction of enlarging the data availability using more cameras in parallel and consequently more processing modules for providing the previously discussed steps (Fig. 4) . In addition another important task is:

5. Data fusion: *the task of fusing information coming from different sensor sources to provide better detection* *and tracking, or to enlarge and enrich the recognition and understanding.*

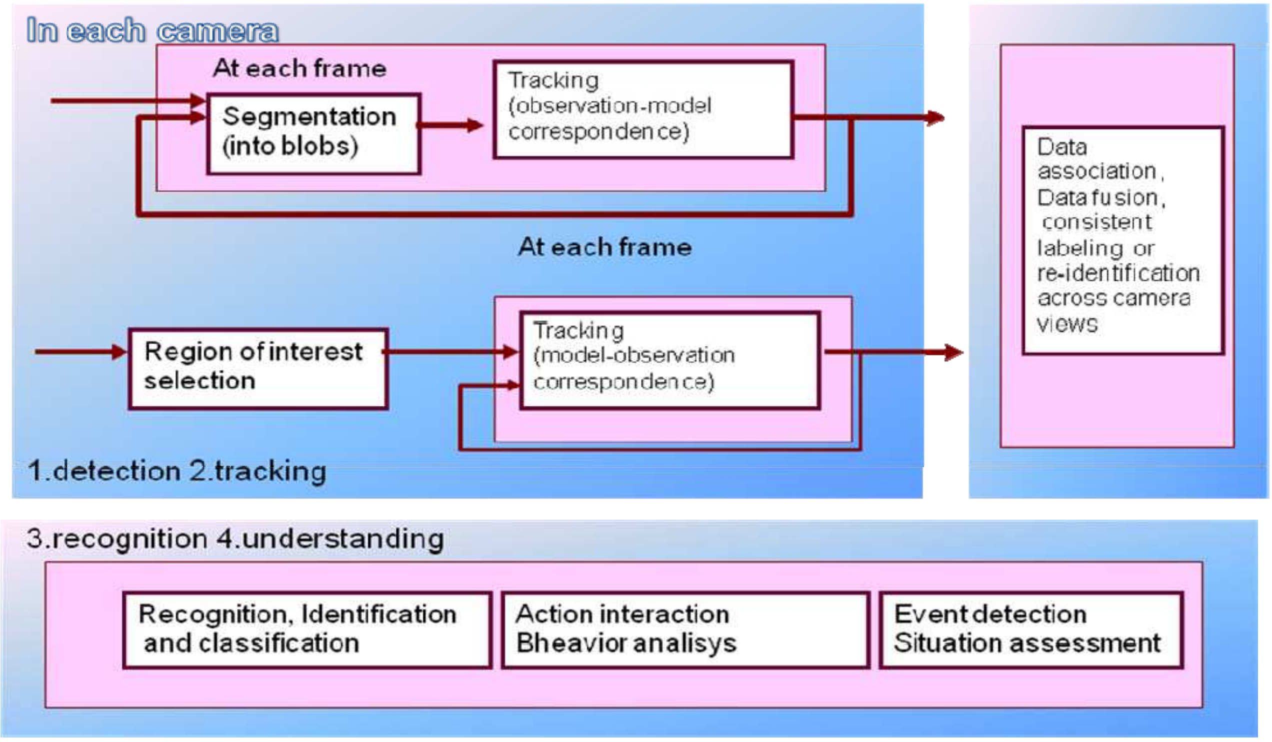

Fig. 4 Typical processes in surveillance systems

From the point of view of the architecture surveillance exhibits the same trends of computing systems, from single to parallel machines. All the issues connected to synchronization problems, processing delocalization, resource sharing, topology and communication of parallel architecture must be taken into account.

We may suggest a Pindaric fly to find a symmetry with the taxonomy of computer architectures, in order to analyze the panorama of new surveillance systems. We can divide them in two categories.

3.1 MULTI-CAMERA SURVEILLANCE SYSTEMS:

Multi-camera (or multi-view) Surveillance Systems are the one based on multiple cameras with overlapping or partially overlapping FoVs (Fields of View) with a fully synchronized acquisition to work on a single enlarged and fused visual data stream. They are typically constituted of fixed or PTZ cameras connected with a single frame grabber in a single server. Here different visual data are acquired, pre-processed and correlated to create an enlarged view and an enriched knowledge source: thus (Fig. 5) many processes P work concurrently on own local memory data M and must be synchronized to produce a shared visual data used by higher level processes, in a producer-consumer paradigm. This model has a direct correspondence with *shared-memory architecture* model, and inherits all the characteristics and requirements well known in tightly coupled multiprocessors. Low Level processes, viewed as threads working on the same processor or on different cores, execute initial steps of surveillance and reconstruct a shared scene and shared visual data where higher level processes can work.

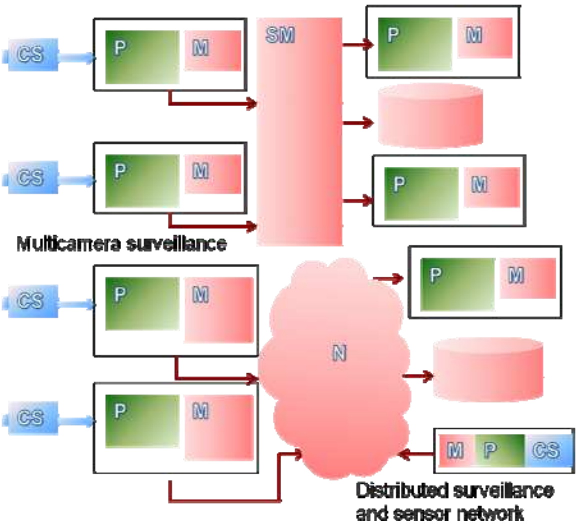

Fig. 5 Multi-camera and distributed surveillance

A necessary task in multi-camera system is *consistent labeling* of tracked objects in the space with overlapping FoVs, generally achieved with *geometric and statistical approaches*. In the ImageLab solution [13] each thread, separately provides segmentation, tracking and people recognition. In the offline process (Fig. 6) using a person as a probe (first row) a task of data fusion allows to create the Entry Edge of Field of Views (second row) and to create the shared space (third row). Each person is detected and tracked in the image space then the information extracted are correlated in a shared space where the homographic reconstruction of the scene is performed. Using epipolar lines and homography, data are fused and the same identifier is given to the same object detected in multiple views, allowing a good disambiguation in the case of occlusions.

There are the same pros and cons of shared memory architecture: basically, by having all data in a shared space high level threads can exploit all information at pixel level or in an intermediate level (e.g. the moving blobs); higher accuracy can be provided and also a 3D reconstruction is possible. On the other hand, the architecture is less scalable and the capability is bounded by the capacity of the system to work in a multi-thread fashion with many streams in parallel. Many communication paradigms can be supported in the shared space: as the most immediate producer-consumer stream we could think also to a cooperative workload where a process working on a view could ask to another process working on another view to have more information for recognition or reasoning.

Another example is the master-slave paradigm which exploits static and PTZ camera to focus on some target such as a face of an individual. This model can be designed in a shared memory space if both modules works on pixels acquired and correlated with geometric transformation to pass from the image plane to a camera to the image plane to another [25].

3.2 DISTRIBUTED SURVEILLANCE SYSTEMS

Distributed Surveillance Systems are based on multiple camera and processing module without the need of overlapping FoVs to fuse surveillance data and information in an enlarged surveillance space. They are constituted of fixed, moving, PTZ cameras in a loosely-coupled architecture where each node is in charge on the own (P-M) pair as much as the process is possible; then connection with higher level data fusion models or other sensor modules is exploited through a network as in a *message-passing architecture*.

Fig. 6 some results of HECOL [13] system for multi-camera surveillance

This architecture is potentially more flexible since the nodes can have a more or less powerful processing capability. A node can be either a smart camera [26] or a classical single stationary surveillance system, either a multi-camera system with a possibly infinite scalability. Moreover also heterogeneous sensors can be easily integrated. Here the challenges are similar to the ones of parallel distributed architectures, defining the most profitable tradeoff between local and remote computation, and computational power; the network bandwidth may be the obvious bottleneck so that normally the exchanged messages regard only high level knowledge (e.g., a people ID or textual data) and, only if necessary, the related snapshot [27].

By working in a distributed environment, also the algorithms are different: for instance the problem of maintaining the same identity of the individual along the time, previously called consistent labeling, here is normally named *re-identification* and is typically addressed with *similarity search* as in the multimedia retrieval. Many works have been proposed to identify people recognized on distributed systems [28]. Basically, the higher level reasoning, by starting from a query coming from a module where an entering person is detected, asks to other modules to give the visual information of the people detected previously [29]. The same paradigms are exploited in non real-time security systems used for instance for forensics application where the query are done on a distributed database.[30]. Instead in surveillance real-time requirement is important and the system should provide high reactivity: a recent research in distributed surveillance in Imagelab [31] exploits the “pathnodes” model to create virtual connection between views. Each module processes each data and tracks people with particle filtering; when a person exits from a FoV his/her particles are propagated along the pathnodes to other modules where the data association starts to search if a person with similar visual features (e.g., texture or color) is detected. Similar approaches exploit distributed mobile agents [32].

3.3. DISTRIBUTED SURVEILLANCE AND SENSOR NETWORK

The distributed paradigm can be extended to the case of sensor networks. Sensors can be visual sensors or based on other sensory data such as RFID, or PIR to infer object presence, identification or position.

They are characterized by a massively distributed architecture, with typically a limited computation capability and a wireless connection. As in the case of *ubiquitous architectures,* they can overcome these limitation by exploited their miniaturization, low power and easy installation features. As in ubiquitous architectures, the challenges are related to node communication, node topology, node replacement and reliability and recovery after failure.

Often they are equipped with embedded processors and FPGAs with sufficient power to provide most of the task required. An example is the MeshEye [33] developed at Stanford for distributed surveillance.

Although the motes have often a limited floating point capability, they can provide most of the simple processing tasks: for instance they can achieve motion detection and communicate to other nodes only frames or region of interests; they can detect moving object, possibly classify people and vehicle and sending to the server nodes only the region of interest within the frame, or they can also detect some very specific behaviors (e.g., an object stopping in a forbidden area) and communicating it to the server. The capabilities of sensor networks will allow to replace expensive system for many distributed tasks[34]

Their architecture allows an use in a moving environment as a wearable computing or mounted on vehicles. This is one of the most emerging technology which will increase in his use in the next future since it will be provide d by similar power than other architecture and is potentially the most flexible and scalable.

4. DESIGN ISSUES IN DISTRIBUTED SYSTEMS

In this last session we would like to underline some design issues which must be taken into account in the next generation of Wide Area Surveillance and Ubiquitous Surveillance systems. These aspects are attached in many current research projects and are well addressed in the literature. Thus, we will provide some reference of interesting results regarding

a) sensor topology
b) architecture topology
c) communication aspects
d) data fusion levels
e) data processing.

In wide area surveillance the definition of the best sensor topology is critical for both multi-camera and distributed systems. In [35] the best placement definition using linear programming and some heuristics is proposed for indoor surveillance, while in [36] same results are achieved using simulating annealing. The authors of [37] present the definition of a 3D Visibility model of a tag (for people) and a suitable optimization via binary integer programming to place in the best way the cameras in a smart room to detect the parts of the human body.

In [38] Kankanhalli et al. proposed a model based on game theory to define the best placement of sensors for both people and vehicles.

A dual problem instead is to learn the relationships between cameras FoVs for learning the best placement. Whil in [39] we proposed to learn the geometric relationships in a multi-camera system, Ellis et al. [40, 41] learned the connection between camera FoVs in distributed system detecting the topology of an existing camera network. Similarly, in [42] the topology of a camera network is inferred by measuring statistical dependence between entrances and exits with a GPS-based performance analysis. A recent overview of the problem related with camera placement in Wide area Network is [43].

A second aspect is *architecture topology*, since, also in the case of a distributed system, the processing module architecture and their connection can be very different: the architecture can be

- <u>Centralized</u>, as in PRISMATICA project [44]: with central servers for high level processing connected with dedicated nodes for cameras, audio, smart cameras and so on;
- <u>Semi-distributed</u>, as in ADVISOR project [45]: many independent nodes each one connected with more cameras, so that each one is a multi-camera system;
- <u>Distributed network with embedded systems</u> [46], a typical distributed architecture with message passing
- <u>Distributed network with agent-based communication</u> [47].

A third problem is related with *communication aspects*; in multi-camera systems the problems are relative to the shared memory and the synchronization, while in distributed architecture the biggest challenges are the bandwidth allocation [48] and the wireless and mobile protocols [49]. Moreover, in the field of sensor network, a good survey is [50].

In all distributed surveillance but especially in sensor network problems of data security and privacy must be taken into account. The work of Perrig et al. [51] address aspects of security in sensor network for surveillance too.

A further aspect is the data fusion level, that is the granularity of data which must be transferred to be fused. Previously, we discussed some of these issues for camera-based systems but new surveillance systems include other different sensors and the fusion of heterogeneous information is more challenging. Examples of research activities are solutions with color and thermal cameras [52], with mobile, wireless, fire, and sound [53] and with visual and PIR sensors [54]. While RFID can give the information about the identity of people wearing RFID tags, a PIR sensor does not receive identity but can localize more precisely the presence of a person which can be used either as a trigger for a video surveillance, or to implement virtual tracking when objects are not visible by cameras.

The last but not less important aspect is related to the *data processing*, that is which algorithms of computer vision and pattern recognition are more suitable. Some discussion has been presented for single camera, multi-camera and distributed cameras, but the problem is also related with the possibility to measure the performance of data processing. The recent work of Chellappa et al. [55] poses an interesting question: actually, in real contexts, noise, errors in homography, lack of planar constraints introduce uncertainty in the position so that the data association, the consistent labeling or re-identification may become very imprecise. In this paper, they propose a model to define the uncertainty of the model and measure it. Finally, it is important to remind the role of benchmarking to measure the performance of multi-camera and distributed systems. Many new benchmarks have been presented, starting from PETS series, ILids, CAVIAR and many others. In our experiments we use these data and others we made available on Visor project www.openvisor.org containing hundreds of hours of videos for research purposes only, with the suitable XML-based annotation, a defined ontology and tools, based on VIPER for performance evaluation. Details are in [56].

Most of the aspects here discussed are deeply covered in a survey of some years ago [57] and in a recent research book [58].

5. CONCLUSIONS

This paper presents a small survey of the issues related with designing smart surveillance systems with multiple tightly or loosely coupled sensors or sensor networks. This research field is ever growing and must cope with hardware, architectural and algorithmic problems.

A possible soft classification is proposed and several aspects have been pointed out. This work is partially supported by BESAFE Nato Project.

REFERENCES

[1] R.Strzodka, C. Garbe Real-Time Motion Estimation and Visualization on Graphics Cards Proceedings IEEE Visualization 2004, pages 545–552, 2004.

[2] G. Gualdi, A. Prati, R. Cucchiara, "Video Streaming for Mobile Video Surveillance" in IEEE Transactions on Multimedia, pp. 1142-1154, October, 2008

[3] D. Koller, J. Weber, T. Huang, J. Malik, Towards Robust Automatic Traffic Scene Analysis in Real-Time," in Proc of ICPR, 1994

[4] V. Kastrinaki , M. Zervakis , K. Kalaitzakis A survey of video processing techniques for traffic applications Image and Vision Computing 2003

[5] Haritaoglu, D. Harwood, and L.S. Davis, "W4: real-time surveillance of people and their activities," IEEE Trans. PAMI, (22) 8, 2000.

[6] R. Cucchiara, C. Grana, M. Piccardi, A. Prati "Detecting Moving Objects, Ghosts and Shadows in Video Streams", IEEE Trans on PAMI, 2003

[7] M. Shah, O. Javed, K. Shafigue Automated Visual Surveillance in Realistic Scenarios IEEE MULTIMEDIA 2007

[8] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. PAMI, (22) 8, 2000.

[9] Tao Sawneir, Kumar. Object tracking with bayesian estimation of dynamic layer representation IEEE Trans on PAMI 24,1 2002

[10] A.Elgammal, R.Duraiswami, D.Harwood, L.S. Davis Background and Foreground Modeling Using Nonparametric Kernel Density for Visual Surveillance Proceedings of the IEEE 2003

[11] Mykhaylo Andriluka, Stefan Roth, Bernt Schiele, "People-tracking-by-detection and people-detection-by-tracking," cvpr, pp.1-8, 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008

[12] Tao Zhao, Ramakant Nevatia, Tracking Multiple Humans in Complex Situations, IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, n. 9, 2004, pp. 1208-1221

[13] S. Calderara, R. Cucchiara, A. Prati, "Bayesian-competitive Consistent Labeling for People Surveillance" in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, n. 2, pp. 354-360, 2008

[14] Dorin Comaniciu, Visvanathan Ramesh: Mean Shift and Optimal Prediction for Efficient Object Tracking. ICIP 2000

[15] M. Isard and A. Blake. A smoothing filter for condensation. In Proc. ECCV, 1998.

[16] Wu, H; Sankaranarayanan, A; Chellappa, R; Online Empirical Evaluation of Tracking Algorithms, IEEE Trans. PAMI 2009 in press

[17] Moeslund , Hilton, Kruger A survey of advanced in vision-based human motion capture and analysis Computer Vision and Image Understanding vol 104 2006

[18] Yilmaz Javed Shah Object tracking a survey ACM Computing Survey vol 33 n 4 2006

[19] Christian Wojek1, Bernt Schiele. "Performance Evaluation of Single and Multi-feature People Detection", Pattern Recognition, Volume 5096/2008,

[20] L. Chen,R. Feris, Y Zhai, L Brown, A Hampapur An Integrated System for Moving Object Classification in Surveillance Videos AVSS2008

[21] J. Aggrawal, S.-Park Human motion: Modeling and recognition of actions and interactions Second Symposium on 3D Data Processing, Visualization and Transmission (3DPVT'04)

[22] G. Gualdi, A. Prati, R. Cucchiara, "Covariance Descriptors on Moving Regions for Human Detection in Very Complex Outdoor Scenes" in Proceedings of ACM/IEEE International Conference on Distributed Smart Cameras (ACM/IEEE ICDSC 2009), Como, Italy, 30-2, 2009

[23] A. Prati, S. Calderara, R. Cucchiara, "Using Circular Statistics for Trajectory Analysis" in Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, Alaska (USA), June 24-26, 2008

[24] Alexei Gritai, Yaser Sheikh, Cen Rao, Mubarak Shah: Matching Trajectories of Anatomical Landmarks Under Viewpoint, Anthropometric and Temporal Transforms. International Journal of Computer Vision 84(3): 325-343 (2009)

[25] Alberto Del Bimbo, Giuseppe Lisanti, Federico Pernici, "Scale Invariant 3D Multi-Person Tracking using a Base Set of Bundle Adjusted Visual Landmarks", Workshop on Visual Surveillance (VS), 2009

[26] M Bramberger, A Doblander, A Maier, Distributed embedded smart cameras for surveillance applications Computer 2006

[27]A C. Sankaranarayanan R Patro,2P Turaga, A Varshney,and R Chellappa Modeling and Visualization of Human Activities for Multicamera NetworksEURASIP Journal on Image and Video Processing
Volume 2009

[28] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu and Bruno Ste
PERSON RE-IDENTIFICATION IN MULTI-CAMERA SYSTEM
BY SIGNATURE BASED ON INTEREST POINT DESCRIPTORS
COLLECTED ON SHORT VIDEO SEQUENCES ICSCD2008

[29] Michel Lantagne, Marc Parizeau and Robert Bergevin, "VIP: Vision tool for comparing Images of People", Vision Interface 2003

[30] Paul Koppen and Marcel Worring. "Multi-target Tracking in Time-lapse Video Forensics", ACM Workshop on Multimedia in Forensics (MiFOR), 2009

[31] R. Vezzani, D. Baltieri, R. Cucchiara, "Pathnodes integration of standalone Particle Filters for people tracking on distributed surveillance systems" in Proceedings of 25th International Conference on Image Analysis and Processing (ICIAP2009), Vietri sul Mare, Salerno, Italy, September, 8-11, 2009

[32]Q. Wu, N. Rao et al On Computing Mobile Agent Routes for Data Fusion in Distributed Sensor Networks ieee Trans on Knowledge and Data engineering vol 16 2004

[33] S. Hengstler D. Prahanth, S. Fong,and H. Aghajan "MeshEye: a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance", Proceedings of the 6th international conference on Information processing in sensor networks, 2007

[34]M Valera, SA Velastin - Intelligent distributed surveillance systems: a review IEEE Proceedings Vision, Image and Signal Processing 2005 -

[35] E. Hoester, R. Linheart Optimal Placement of Multiple sensors in Multicamera network principles and application. Academic Press 2009, proc of ACM VSSN 2006

[36] J. Mittal, L. Davis A general method for sensor planning in multi-sensory systems: extension to random occlusion, Journal of Computer Vision 2008

[37] J. Zhao S. Cheung, TNnguyen ,MULTICAMERA SURVEILLANCE WITH VISUAL TAGGING AND GENERIC CAMERA PLACEMENT Multicamera network principles and application. Academic Press 2009

[38] Singh, V.K.; Kankanhalli, M.S.; "Adversary Aware Surveillance Systems", IEEE Transactions on Information Forensics and Security, Volume 4, Issue 3, Sept. 2009 Page(s):552 – 563

[39] S. Calderara, A. Prati, R. Cucchiara, "HECOL: Homography and Epipolar-based Consistent Labeling for Outdoor Park Surveillance" in Computer Vision and Image Understanding, vol. 111, n. 1, pp. 21-42, 2008

[40] T.Ellis, D. Makris, J.Black Learning a multicamera topology Proc of VS& PETS2003

[41] T.Ellis, D. Makris, J.Black Bridging the gap across cameras CVPR2004

[42] Tieu, Dalley, Grimson Inference of non overlapping camera network topology by measurng statistical dependence Proc of ICCV 2005

[43] Besma R. Abidi, Nash R. Aragam, Yi Yao, Mongi A. Abidi: Survey and analysis of multimodal sensor

planning and integration for wide area surveillance ACM Computing Reviews 2009

[44] P. Lai Lo, J. Sun, s. Velastin Fusing visual and audio information in a distributed surveillance system ACTA Automatica SINCA 2003

[45] M. Valera Espina and S.A. Velastin, "Intelligent Distributed Surveillance Systems: A Review," IEE Proc. Vision, Image and Signal Processing, Apr. 2005, pp. 192—204

[46] M. Christensen, R. Alblas V2 design issues in distributed video surveillance systems Denemark 2000

[47] M. A. Patricio, J. Carbó, O. Pérez, J. García, and J. M. Molina Multi-Agent Framework in Visual Sensor Networks EURASIP Journal on Advances in Signal Processing Volume 2007 (2007)

[48] C. Regazzoni, V.Ramesh, G. Foresti Special Issue on video communication processing and understanding for third generation surveillance systems Proc. Of IEEE 2001

[49] F. Akyildiz, W. Su, , Y. Sankarasubramaniam and E. Cayirci Wireless sensor networks: a survey Computer Networks, Volume 38, Issue 4, 15 March 2002

[50] M Tubaishat, S Madria Sensor networks: an overview - IEEE potentials, 2003

[51] Adrian Perrig , J.Stankovic , D. Wagner, “Security in wireless sensor networks” Communications of the ACM Volume 47, Issue 6 (June 2004)

[52] J.Han, B.Bhanu Fusion of color and infrared video for moving human detection Pattern recognition 2007

[53] Y Tseng, Y Wang, K Cheng, Y Hsieh Imouse: an integrated mobile surveillance and wireless sensor system IEEE computer 2007

[54] R. Cucchiara, A. Prati, R. Vezzani, L. Benini, E. Farella, P. Zappi, "Using a Wireless Sensor Network to Enhance Video Surveillance" in Journal of Ubiquitous Computing and Intelligence (JUCI), vol. 1, pp. 1-11, 2006

[55] A. C. Sankaranarayanan, A.Veeraraghavan, and R.Chellappa, Object Detection, Tracking and Recognition for Multiple Smart Camera. Proceedings of the IEEE | Vol. 96, No. 10, October 2008

[56] R. Vezzani, R. Cucchiara, "Video Surveillance Online Repository (ViSOR): an integrated framework" (available online) in Multimedia Tools and Applications, DOI 10.1007/s11042-009-0402-9, 2009

[57] Hu, Tan, Wang, Maybank: A Survey on visual surveillance of object motion and behaviors. IEEE Trans. On System Man and Cybernetics, vol. 34 n. 3 2004

[58] Multicamera network principles and application Aghajan, Cavallaro eds. Academic Press 2009

Swarming Machines & Sensor Fusion
What can we learn from biology?

Verena V. Hafner
Cognitive Robotics, Dept. of Computer Science, Humboldt University Berlin, Germany

In robotics, strategies to perceive the environment and make sense of the information coming from different sensory modalities are an important topic. Even more so in recent years, when the traditional approach of sense-think-act got replaced by looking at the complex sensorimotor interaction of a physical agent with its environment and with other agents.

A typical situation that is interesting for studying the phenomena of embodied artificial intelligence and one that is also relevant to many applications is navigation behaviour. How is it possible to find one‘s way back to a certain place, which information need to be processed and which information need to be stored and remembered? This not only depends on the available sensory information, but also on the goal and the particular strategy applied to reach this goal. Navigation strategies such as visual homing have shown that it not always necessary to know one‘s own spatial position or relationship to other places in order to navigate to a given goal.

In these situations, strategies that merge information on a higher level have to be developed and combined with methods of low-level sensor fusion. By extending the situation from a single agent interacting in its environment to a swarm of agents, the question of which information to process and to share becomes even more prominent. Simple sensorimotor rules result in complex emergent behaviour. Swarm animals succeed in solving the problem of sensor fusion with limited sensory and computational capacities. Extracting and understanding these principles will help us to develop new computational methods.

In this presentation, I will discuss these concepts and ideas and show how they could be applied to a swarm of flying robots.