Privacy Sensitive Surveillance for Assisted Living – A Smart Camera Approach

Sven Fleck and Wolfgang Straßer

1 Introduction

An elderly woman wanders about aimlessly in a home for assisted living. Suddenly, she collapses on the floor of a lonesome hallway. Usually it can take over two hours until a night nurse passes this spot on her next inspection round. But in this case she is already on site after two minutes, ready to help. She has received an alert message on her beeper: "Inhabitant fallen in hallway 2b". The source: the SmartSurv distributed network of smart cameras for automated and privacy respecting video analysis. Welcome to the future of smart surveillance! Although this scenario is not yet daily practice, it shall make clear how such systems will impact the safety of the elderly without the privacy intrusion of traditional video surveillance systems.

The demand for video surveillance systems is immensely increasing in various fields of security applications and beyond. Camera systems cover, e.g., airports, train stations and are nowadays also installed in trains and some aircrafts. Further examples include monitoring of power plants, shopping malls, museums, parking garages and hotels. A typical casino in Las Vegas uses over 2,000 cameras [39, 11]. Over 4 million cameras are installed in the UK. Moreover, additional applications arise that are not motivated by security but have its purpose in marketing, e.g., statistics in retail stores like measuring shelf attractiveness or getting information about the customers (gender, child/adult etc.). Traffic monitoring is another field. Areas aiming at increased individual safety independent of terrorist threats further extend the application range of such systems. One important example is the field of assisted living where we currently concentrate on. According to [24] the US population over

Sven Fleck
SmartSurv Vision Systems GmbH, spin-off of University of Tübingen, WSI/GRIS, Tübingen, Germany, e-mail: `fleck@smartsurv.de`

Wolfgang Straßer
University of Tübingen, WSI/GRIS, Tübingen, Germany, e-mail: `strasser@gris.uni-tuebingen.de`

H. Nakashima et al. (eds.), *Handbook of Ambient Intelligence and Smart Environments*,
DOI 10.1007/978-0-387-93808-0_37,

age 85 will triple from 4.5 million in 2003 to 14.2 million by 2040. Automated 24/7 surveillance to ensure safety of the elderly while respecting privacy becomes a major challenge. Falls are one of the most serious health risks among seniors over the age of 65 [30], affecting more people than stroke and heart attacks combined. Moreover, detecting falls to get immediate help reduces the risk of hospitalization by 26% and death by over 80%.

Traditional "CCTV"[1] surveillance systems used in the field with their centralized processing and recording architecture together with a simple multi-monitor visualization of the raw video streams bear several drawbacks and limitations. The most relevant problem is the total lack of privacy. Additionally, the necessary communication bandwidth to each camera and the computational requirements on the centralized servers strongly limit such systems in terms of expandability, installation size and spatial & temporal resolution of each camera. As the visualization is counterintuitive and fatiguing due to the massive load of raw video data, manual tracking is not always performed reliably and activities often remain undetected. For example, tracking a suspect on his way requires to manually follow him within a certain camera and when he leaves one camera's view, an appropriate camera has to be identified, manually selected. Furthermore, the operator then has to put himself in the new point of view to keep up tracking. Obviously, this already strongly limits the number of persons concurrently tracked. A military study states that after approximately 22 minutes, an operator will miss up to 95% of scene activities [5]. The personnel costs are another major issue. An automated and privacy respecting 24/7 surveillance system is thus a strongly desired goal. Latest video analysis systems emerging today are based on centralized approaches which impose strong limitations in terms of expandability and privacy.

A ***distributed smart camera approach*** is proposed that instead qualifies for surveillance in privacy sensitive areas due to the fact that major processing is performed ***automated*** and ***embedded*** in each camera node. A real world prototype system at a home for **assisted living** has been set up. The proposed system's goal set is to analyze the real world and reflect all relevant – and only relevant – information in an integrated virtual counterpart, live and in a 24/7 way. Important is that privacy recommendations of the specific application are taken into account. The top level overview of the proposed concept and system is illustrated in Fig. 1. It is a fully sensor driven system.

Section 2 gives a classification of video surveillance systems. After describing the related work (Section 3), Section 4 describes the problems of centralized and distributed surveillance systems. In Section 5 the architecture of our system is presented, the smart camera component is detailed in Section 6. Section 7 presents the visualization component and at the same time illustrates results of the whole system, followed by the conclusion.

[1] *CCTV* – Closed Circuit Television

Fig. 1 Overview of proposed approach for smart surveillance.

2 Classification of AAL & Video Surveillance Systems

Video surveillance systems can be classified by the following criteria:

- **Manual vs. automated** – This defines if the video analysis is performed by a human operator or if relevant information is already derived from the raw video stream in an automated way.
- **Centralized vs. distributed** – Automated analysis can be further divided by two philosophies, if the processing is performed on a centralized server installation or distributed and embedded within a smart cameras network.

Fig. 2 illustrates where the proposed – distributed & automated – approach is placed within the design space of camera technology and degree of automation. The camera technology is denoted on the abscissa and can be divided into centralized and decentralized approaches. On the ordinate the degree of automation (and thus the level of abstraction of the results) in video scene analysis is given. This starts from a purely manual (CCTV) approach, continues over motion detection to video analysis in image domain. Although this already includes the abstraction from pixel level to object level, it is still tied to individual cameras. This in combination with a centralized camera approach is the typical combination of automated state of the art systems today. Moving such approach to smart cameras would lead to the blue circular area on the bottom right. Although this is then automated and executed in a distributed way, not the full potential of smart cameras is leveraged: All cameras act as individual cameras with no common world model.

The approach in this article instead addresses tracking and activity recognition in world coordinates and thus to abstract from camera domain to world domain. This in combination with a smart camera based approach for distributed processing is depicted as yellow circle.

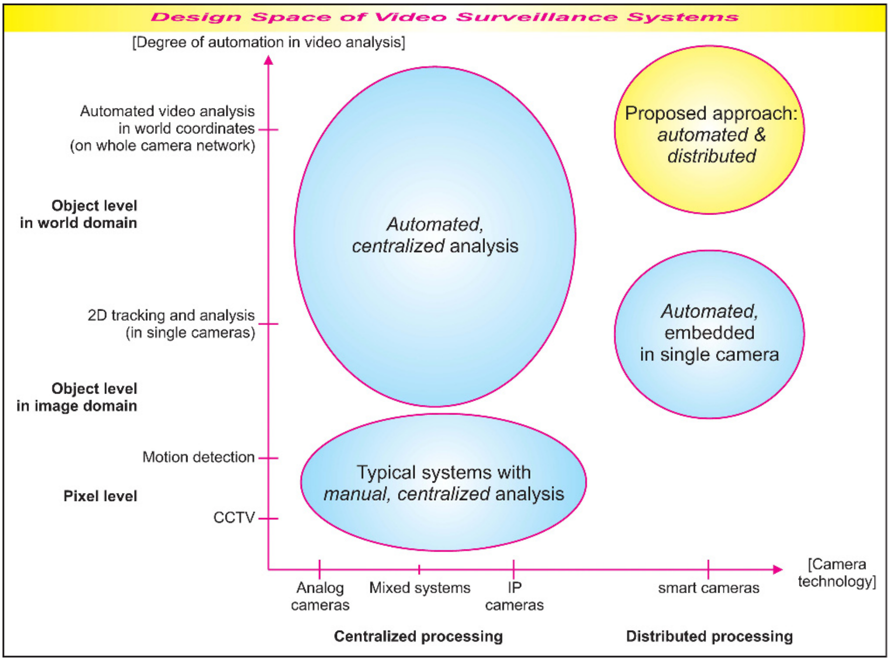

Fig. 2 Classification of video surveillance systems with respect to degree of automation and camera technology.

3 Related Work

3.1 Related Work – Surveillance

Traditional CCTV systems obviously belong to the class "*manual*" where all raw video data is analyzed by human operators and centrally stored using DVR[1] solutions.

3.1.1 Automated Surveillance Systems

Yilmaz, Javed and Shah give an extensive survey on object tracking [48]. The IEEE Signal Processing issue on surveillance surveys the current status of automated surveillance systems, e.g., Foresti et al. present "active video based surveillance systems". Siebel et al. especially deal with the problem of multi camera tracking and person handover within the ADVISOR surveillance system [38]. Hampur et al. describe IBM's S3 comprehensive multiscale tracking system which is also detailed in [23] and includes distance estimation using stereo vision. The IE book on "Intelli-

[1] *DVR* – digital video recording

gent distributed video surveillance systems" [43] also provides a broad overview. At UCSD, Trivedi et al. presented distributed interactive video arrays (DIVA) for situation awareness [42]. This work encompasses a comprehensive system for traffic and incident monitoring [34] which has been installed at multiple real world scenarios around San Diego (I-5, Qualcomm stadium, Gaslamp quarter, Coronado bridge). Shah et al. presented Knight, an automated real world surveillance system [37] running on a centralized server. It mainly targets railroad crossing scenarios. A complete system for aircraft activity monitoring within the AVITRACK project is described by Kampel et al. in [27]. Collins et al. in collaboration with DARPA show the implementation of a system for autonomous video surveillance in [6]. A broader overview about human activity recognition is given in [33] by Ribeiro et al.

All these systems share the fact that they are host *centralized* which requires the transmission of video feeds which costs bandwidth and leaves the door wide open for misuse. A *distributed* smart camera approach instead offers many benefits, only very few systems have been published that take advantage of those. Quaritsch et al. presented a multi camera tracking method in image domain on distributed, embedded smart cameras with a fully decentralized handover procedure between adjacent cameras [32].

3.1.2 Privacy Respecting Surveillance

Actually relatively few surveillance systems addressing privacy issues have been published: Schiff et al. proposed a very interesting approach, the *"respectful camera"* [35] where the problem of automatically obscuring faces in real time to assist in visual privacy enforcement is considered. To this end, persons are required to wear colored hats or vests. This approach primarily targets surveillance applications where the raw video stream is necessary, e.g., if an alarm is detected. Also at CVPR, a workshop on privacy research in vision has been started where research is taken out to scramble videos as shown by Dufaux and Ebrahimi [14] or de-identificate faces as presented by Gross et al. [22]. A layered approach where privacy levels are coupled to users' access control rights is presented by Senior et al. [36]. Cucchiara et al. at the University of Modena also cover face obscuration [9] within their comprehensive video surveillance approach.

3.1.3 Assisted Living

Assisted living systems can be divided into *manual* and *automated* ones. State of the art products like Philips Lifeline [30] are manual systems, i.e., the user wears a wireless help button which he can press in case of an emergency, e.g., in case of a fall – if he is still in the condition of doing so. This already tremendously helps. Automated systems go further and can be classified into *non-visual* and *visual* systems. Intel's Proactive Health Strategic Research Project [13] and Tabar et al. [40] belong to the former class. Each senior has to wear a tag which however can be lost or forgotten easily. Visual surveillance by stationary cameras instead provides contactless observation without the need to equip the elderly. Important to note is the work by Aghajan et al., especially [1, 28] where a distributed accident management system for assisted living is presented which combines a wireless sensor solution worn by each person with a visual surveillance system. Privacy is addressed by not recording the camera feed if no accident is detected. Cucchiara et al. [10], Hauptmann et al. [24] and Williams et al. [46] present their approaches for assisted living and visual falling person detection, Toreyin et al. in [41] combine visual cues with audio cues to increase robustness.

3.2 Related Work – Smart Cameras

A variety of smart camera architectures designed in academia and industry exist today and typically target machine vision and automation applications, although recently they start being used in surveillance. What all smart cameras share is the combination of a sensor, an embedded processing unit and a connection, which is nowadays often a network unit. The processing means can be roughly classified in DSPs, general purpose processors, FPGAs and a combination thereof. Wolf et al. identified smart camera design as a leading-edge application for embedded systems research [47]. The special issue [12] of the EURASIP Journal on Embedded Systems provides a comprehensive overview of current research in the field of embedded vision systems. Bernhard Rinner's group designed an innovative smart camera (Bramberger et al. [3]) for distributed embedded surveillance together with ARC. It consists of a network processor and a variable number of DSPs. Chalimbaud and Berry presented a smart camera based on FPGAs [4]. Kleihorst et al. presented a smart camera mote with a Xetal-II high performance yet low power SIMD processor [29]. Hengstler et al. presented MeshEye [25], an embedded, low-resolution stereo vision system on ARM7TDMI RISC microcontroller basis. The idea of having Linux running embedded on the smart camera becomes more and more common (Matrix Vision, Basler, Elphel, FESTO).

From the other side, i.e., the surveillance sector, IP based cameras are emerging where the primary goal is to transmit live video streams to the network by self contained camera units with (often wireless) Ethernet connection and embedded processing that deals with the image acquisition, compression (MJPEG or MPEG4), a webserver and the TCP/IP stack, and offer a plug 'n' play solution. Further processing is typically restricted to, e.g., user definable motion detection. All the underlying computation resources are normally hidden from the user.

The border between the two sides becomes more and more fuzzy, as the machine vision originated smart cameras get (often even GigaBit) Ethernet connection and on the other hand the IP cameras get more computing power and user accessability to the processing resources. For example the processors of the Axis IP cameras are also fully accessible and run Linux.

4 Identified Problems of Current Approaches and Resulting Impact

To summarize, despite the benefits that video surveillance systems provide, the systems used today impose severe problems:

- **Problem #1: Privacy.** Privacy is totally ignored by today's systems that fully rely on the accessability to raw video feeds. Both on the way to the operator and the operator itself impose potential risk points of misuse.
- **Problem #2: Overstraining data.** The sheer mass of raw video data is simply overstraining the operator. After approximately 22 minutes, even a concentrated operator will miss up to 95% of scene activities [5]. Distractions like the zooming to attractive women as documented by studies [26] make it even worse. Tracking and handover even of single persons on camera level is a challenging task, tracking multiple persons concurrently is practically impossible.
- **Problem #3: Limited flexibility & expandability.** Adding new cameras to a centralized system can easily exceed the available bandwidth or the processing capabilities of the centralized server.

To get more insight in why privacy is identified as the **problem #1** of today's centralized surveillance architectures a brief overview of privacy as property is given.

4.1 Privacy

Privacy is a fundamental and very personal property to be respected – also by law – to allow each individual to keep control of the flow of information about himself. The definition of privacy needs to be seen in a nuanced light as it comprises multiple individual points. According to [16], privacy comprises confidentiality, anonymity, self-determination, freedom of expression and control of personal data. In the world of video surveillance it is especially important to guarantee privacy, as persons within a perimeter covered by cameras have very little choice of being surveilled or not, whereas e.g., in the case of cell phone tracking the user still has the choice to turn his phone off.

4.1.1 The Privacy Dilemma and its Impact

The royal academy of engineering gives a nice overview of the dilemmas of privacy and surveillance [16]: "both promise great benefits and pose potential threats. In some cases those benefits and threats are so sharply opposed, and yet so equally matched in importance, that dilemmas arise". One apparent example is visual surveillance within an aircraft to recognize suspicious activities like group forming and attack preparation. On one hand, this could permit the prevention of fatal consequences. On the other this poses an intrusion into privacy of each passenger being monitored from close distance. A second example is within a home for assisted living: monitoring if a person falls while taking a shower or if he falls asleep during taking a bath are extremely helpful. However, the privacy issue is obvious. As no technical solution to this dilemma is known, typically the commensurability between privacy and surveillance is taken by law. Often, the law differentiates if a permanent, 24/7 operation of surveillance takes place, if data is recorded and on the other hand if a notification about the surveillance operation is present and if a special interest (threat) motivates such installations. Thus, only compromises are achieved which limits the added value of surveillance and at the same time still represents a direct intrusion into privacy.

4.1.2 Centralized Systems: Where Does the Privacy Breach Happen?

As shown in Fig. 3 the privacy chain can fail at multiple points: In classical non-automated "CCTV" type of applications security personnel is inherently embedded within the loop to analyze the raw video data. Obviously, this opens the floodgates to misuse (Fig. 3,(1)). Studies report security persons zooming to attractive women and secretly uploading video feeds to publicly accessible portals like YouTube – while at the same time missing relevant activities. Second, both for manual and for centralized & automated systems privacy sensitive video data can be picked up within the transmission channel (even if the signal is encrypted) from the camera all the way to a centralized server (Fig. 3,(2)). Even worse, this can happen unnoticeable, especially if it includes a wireless connection. In both cases no indication is apparent to the person in front of the camera that his privacy relevant video data is misused. With a smart camera based approach only the camera itself remains as

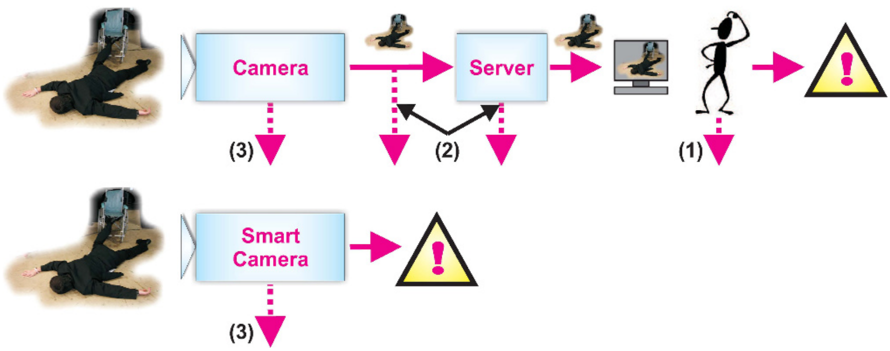

Fig. 3 Privacy breach points in a surveillance system. **Top:** Manual, centralized architecture. **Bottom:** Distributed smart camera based architecture as proposed.

point of attack (Fig. 3,(3)). So, the camera housing should be sealed to prevent both electronical (network attacks) and mechanical access.

5 The Proposed SmartSurv Approach for AAL

Based on [19] the SmartSurv 3D Surveillance System is presented in the following. The system serves as a prototype implementation to come closer to the vision depicted in Section 1.

5.1 The Smart Camera Paradigm

Centralized computer vision systems typically consider cameras only as simple sensors. The processing is performed after transmitting the complete raw sensor stream via a costly and often distance-limited connection to a centralized processing server. We think it is more natural to also physically embody the processing in the camera itself: what algorithmically belongs to the camera is also physically performed in the camera. The idea is to compute the information where it becomes available – directly at the sensor – and transmit only results that are on a higher level of abstraction.

5.2 SmartSurv 3D Surveillance Architecture

The top level architecture of our distributed surveillance and visualization system is given in Fig. 4. It consists of three levels: sensor analysis level (left column), virtual world level (center column) and visualization level (right column). In the following, all components are described briefly.

- **Scene acquisition:** The 3D model acquisition of background scenes with a mobile sensor platform is covered within this component. Alternatively, a 2D floor plan can be used.
- **Scene analysis:** A distributed network of smart cameras for scene analysis where both geo referenced tracking and activity recognition are performed simultaneously, embedded in each camera node. All camera nodes transmit their results to the server level. This smart camera approach allows for optimal scalability, i.e., the processing requirements of a new camera node are completely covered by its embedded processing means. The bandwidth and server performance requirements are only gradually increasing by each camera node. A plug ’n’ play concept allows easy adding of new cameras.
- **Virtual world:** The virtual world acts as an abstraction interface to decouple the visualization clients from the camera based scene analysis. It is implemented by the server level to receive all information from the smart camera network, stores it in the virtual world model and provides the access to this model for visualization clients. It manages configuration and initialization of all camera nodes, collects the resulting tracking and activity recognition data and takes

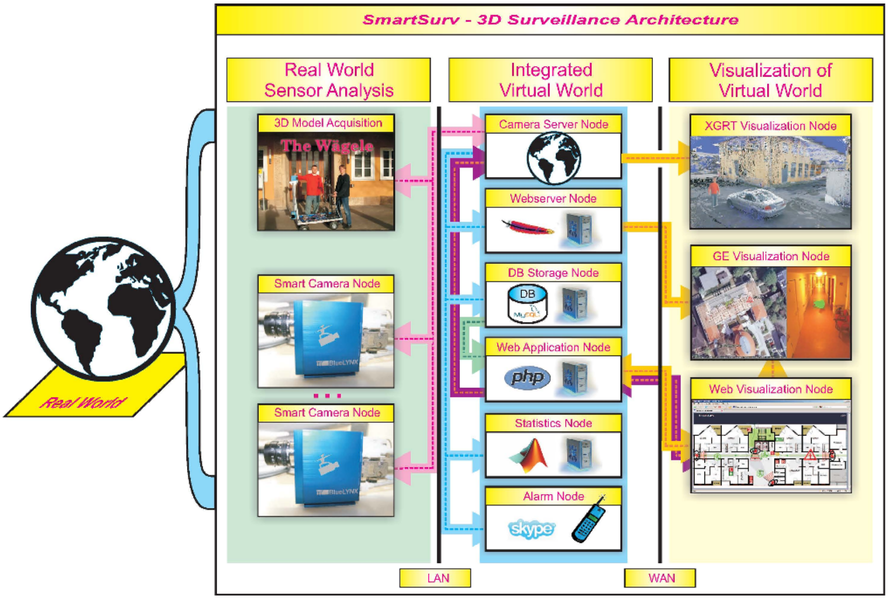

Fig. 4 SmartSurv 3D Surveillance Architecture. **Left column:** Scene acquisition (The Wägele) & distributed scene analysis (smart camera network). **Center column:** Server (virtual world) level. **Right column:** Visualization level.

care of object handover between cameras. It comprises a camera server node as the main component, a webserver and a web application node, a data base storage node, a statistics node and an alarm node. The system's scalability is also ensured on server level. Each node can run on a dedicated server, however, also all server level nodes can run together on a single server PC. Following a Plug 'n' Play concept, a new camera is automatically detected as soon as it is attached to the network and can be added to the server.

- **Integrated visualization:** Relevant results of the distributed scene analysis are embedded live in one consistent and geo referenced 3D world model. Three different visualization options are available.

5.3 Privacy Filters

Depending on the privacy requirements of the application and the used visualization the following options are available:

- The person is represented by an augmented symbol reflecting the position and state (e.g., normal or falling). This is the preferred visualization option.
- The object currently tracked is embedded within the respective visualization. This results in a virtual world model that aims at reflecting as much as possible

the real world. Either the live texture of the ROI covering the tracked object or a silhouette as illustrated in Fig. 15, right can be transmitted.
- The live video feed, e.g., of the most relevant camera can be enabled. An automated switching to display the most relevant camera stream, overlaid with tracking results is available as shown in Fig. 9. For increased privacy demands the silhouette can be transmitted instead as illustrated in Fig. 8.

Future work could use a multi-level privacy filter approach according to the alert level and user privileges, e.g., if a certain medical condition is met, a live video is transmitted, but then also only to the responsible medical doctor.

5.4 Smart Camera Hardware

Smart cameras from Matrix Vision are used within our system. Each smart camera consists of a sensor, a Xilinx Spartan 3 FPGA for low level computations, a PowerPC processor for main computations and an Ethernet networking interface. The mvBlueLYNX 600 series performs its main computations on a 400 MHz Motorola MPC 8245 PowerPC with MMU & FPU running Embedded Linux. It further comprises 64 MB SDRAM (64 Bit, 133 MHz), 32 MB NAND-FLASH (4 MB Linux system files, approx. 40 MB compressed user filesystem) and 4 MB NOR-FLASH (bootloader, kernel, safeboot system, system configuration parameters). IP based connection is used for field upgradeability, parameterization of the system and for transmission of the tracking results during runtime. It consumes about 5 W power. The mvBlueCOUGAR series offers Gigabit Ethernet (GigE) and Power over Ethernet (PoE) as distinctive features besides a more compact form factor (39 x 39 x 96 mm^3).

5.5 SafeSenior - The Installation at An Assisted Living Home

To evaluate the system within a real world scenario, a full prototype system has been installed within a house of Germany's leading provider for assisted living homes for the elderly (see Fig. 5). It comprises four mvBlueLYNX 621CX XGA smart cameras and one server PC. Thereby, the system covers the main hallways of the building's ground level as depicted in Fig. 17 with each cameras' position and field of view marked as stylized eye. Fig. 6 shows the chosen position and viewing angle of the four smart cameras installed in the home for assisted living together with a snapshot image from the viewpoint of each camera. The cameras in the hallways are mounted vertically to better use the spatial resolution of the long hallway. Each camera brings its calibration data with it when registering to the system. It is automatically embedded within the ground plan as shown in Fig. 6, once its camera parameters are known. The system has been running 24/7 for several months now since its installation early 2007 (with no activity recognition at that time). The algorithms have

Fig. 5 Our installation in a home for assisted living. Smart cameras are marked with rectangles.

Fig. 6 Views from the respective smart cameras within the home for assisted living.

been developed with this system as testbed. Although a systematic survey is part of future work, informal questioning of several inhabitants showed that the system is accepted very well. The persons feel safer as their fear of falling is somewhat attenuated. Potential critics due to the fact that cameras are used were very seldom, partly because an information poster clarifies that only automated video analysis takes place with no user access to video data.

6 Smart Camera Architecture

The smart camera architecture is illustrated in Fig. 7. It can be roughly partitioned

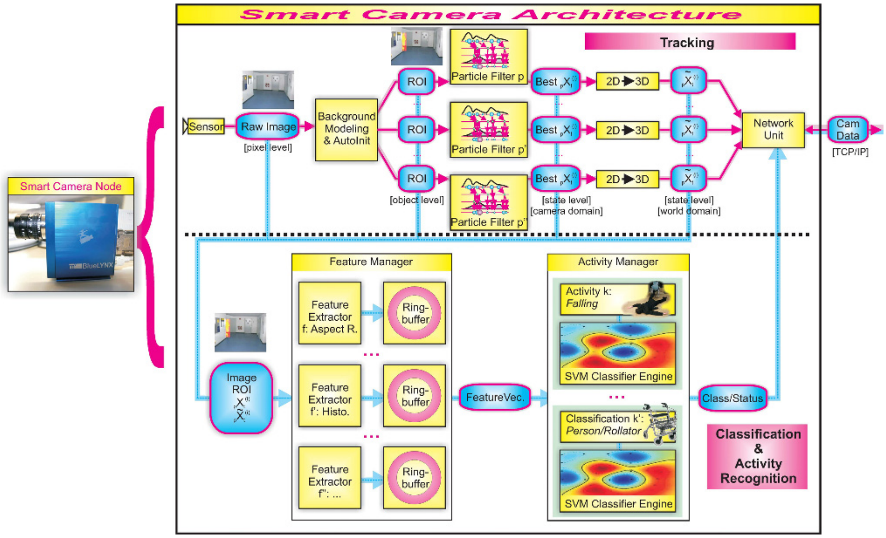

Fig. 7 Smart Camera Node's Architecture. **Top:** Motion detection and tracking. **Bottom:** SVM based activity recognition/classification.

in the tracking related parts (Fig. 7, top) and the classification & activity recognition part (Fig. 7, bottom.) The upper parts consist of the following components which are based on [18] and detailed in the following: a background modeling & auto init unit, multiple instances of a particle filter based tracking unit, $2D \rightarrow 3D$ conversion units and a network unit.

6.1 Background Modeling and AutoInit

Within the **background modeling** unit we take advantage of the fact that each camera is mounted statically. This enables the use of a background model for segmentation of moving objects. This unit has the goal to model the actual background in real time, i.e., foreground objects can be extracted very *robustly*. Additionally, it is important, that the model *adapts* to slow appearance (e.g., illumination) changes of the scene's background. Elgammal et al. [15] give a nice overview of the requirements and possible cues to use within such a unit in the context of surveillance. Due to the embedded nature of our system, the unit has to very computationally efficient to meet the real time demands. Our unit works on a per-pixel basis and is capable of running at 20+ fps on any smart camera used. This is extended by a mixture model to further increase robustness by having multiple background hypothesis alive at the same time. To handle segmentation (i.e., the transformation from raw pixel level to object level), erosion and region growing steps are applied.

6.1.1 Shadow Suppression

Shadow suppression is required to increase the robustness of the applications on top of the tracking, i.e., the activity recognition to work robustly. Prati et al. [31] give an overview of different shadow detection techniques. An existing approach of Cucchiara et al. [7, 8] has been chosen to implement on the smart camera.

6.1.2 Auto Initialization and Destruction

If the region is not already tracked by an existing particle filter, a new filter is instantiated with the current appearance as target and assigned to this region. An existing particle filter that does not find a region of interest near enough over a certain amount of time is deleted. This enables the tracking of multiple objects, where each object is represented by a separate color based particle filter. Two particle filters that are assigned the same region of interest (e.g. two people that walk close to each other after meeting) are detected in a last step and one of them is eliminated if the object does not split up again after a certain amount of time.

6.2 Tracking

Our tracking engine is based on **particle filter** units using color distributions in HSV space. For each person/object p tracked at time t, the state ${}_pX_t^{(i)}$ of hypothesis (sample) i comprises its position, size and velocity. The tracking unit is also capable of adapting the target's appearance during tracking using either the pdf's unimodality (as in [20]) or the background model's information as adaptation rate to circumvent overlearning. On the mvBlueLynx 420CX smart camera (200 MHz PowerPC), we achieve about 17 fps for the whole tracking pipeline when one object is tracked, 15 fps for two objects. The 600 series and the mvBlueCOUGAR (400 MHz) stay above 20 fps, also for multiple objects. Besides the number of objects, i.e., the number of particle filter instances, the frame rate is also affected by the target size on pixel level. We use a subsampling approach to attenuate this effect. Tracking examples are shown in Fig. 8. Please see [18] for further details. The approximated pdf $p({}_pX_t|Z_t)$ for each person/object p tracked is then reduced: only the maximum likelihood sample ${}_pX_t^{(i)}$, representing the most probable position and scale in camera coordinates, is further processed in the $2D \rightarrow 3D$ **conversion unit**.

Fig. 8 Smart camera indoor and outdoor tracking examples.

6.3 $2D \rightarrow 3D$ *Conversion Unit*

Only the maximum likelihood sample, that represents the most probable position and scale in camera coordinates, is further processed in the $2D \rightarrow 3D$ **conversion unit**. Alternatively, the mean can be used. However, in case of multimodal distributions (which particle filters can handle) this would lead to a worse estimate. This unit converts from image domain to a common (and geo referenced) 3D world coordinate system using a flat floor assumption and the camera's calibration data. To summarize, each smart camera node provides its results in geo referenced world coordinates, taking the origin of the local model (e.g., floor plan or Wägele model) into account of which the GPS position is known.

6.4 Handover

To achieve a seamless inter-camera tracking decoupled from each respective sensor node, the **handover unit** merges objects tracked by different camera nodes if they are in spatial proximity. To reduce mismatches during handover due to spatial ambiguities the object's appearance is also taken into account. No video data has to be transmitted for the handover process, a color histogram is used for measuring similarity in appearance. Fig. 9 illustrates a typical handover procedure. This information is also used to track persons beyond the field of view of single cameras. The extracted data is then sent to the server along with the texture of the object (if enabled) using the network unit.

6.5 Activity Recognition

Each smart camera is capable of embedded classification and activity recognition based on a supervised learning approach utilizing Support Vector Machines (SVMs). The camera is trained by showing and manually labeling relevant situations (e.g., falling persons) or objects. All extracted feature vectors are automatically processed in an offline optimization step and the optimized kernel is downloaded onto each smart camera to perform activity recognition for each object currently tracked.

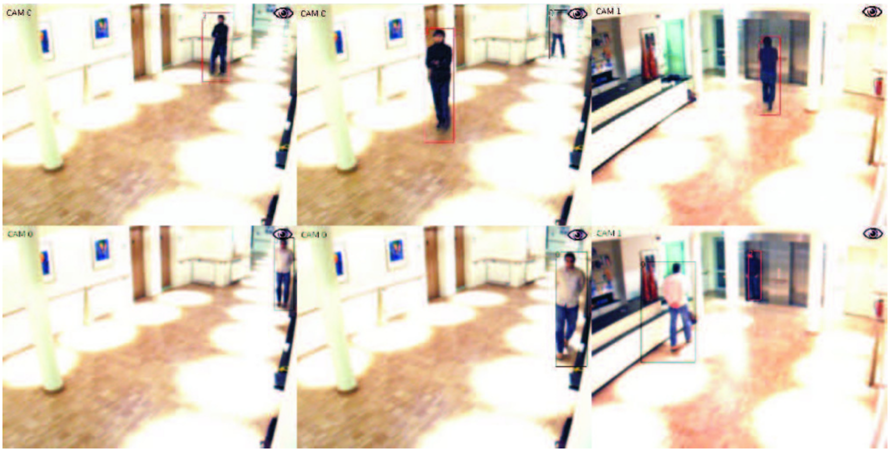

Fig. 9 Automated handover: Each person currently tracked is handed over automatically to the next relevant camera, managed by the server. **Top row:** Shown for the person in dark cloths. **Bottom row:** Shown for the person in the light shirt.

Also object classification is possible with the same framework: The classification works both on static images (e.g., based on appearance) and on their progression over time. The classification & activity recognition unit aims at classifying objects according to their appearance, position (from the tracking unit) or any derived feature vector. The activity recognition comprises two main parts: the **feature manager** and the **activity manager** (see Fig. 7, bottom). Both are described below.

6.5.1 Activity Manager

The **activity manager** holds control of all activities (and objects) k to recognize concurrently. Every activity is instantiated on startup and requests the required feature extractors f from the feature manager which are then sent to the embedded SVM classifier engine. Currently, falling person detection is implemented as a first activity to demonstrate our architecture and at the same time deal with the most important activity with respect to assisted living.

6.5.2 Feature Manager

The **feature manager** collects all requested features of the different activities and builds an intersection of those, so every feature extractor is maximally triggered once per frame. The results of the extractors are then put in a ring buffer, which exists for every extractor. In those ring buffers the results of the last n-frames are hold to represent temporal characteristics. The feature manager then assembles the different feature vectors according to the request of each activity, and returns them to the activity manager. As a first start, within the context of activity recognition five feature extractors are implemented until now with to be used especially for falling person detection. These include the temporal characteristics of aspect ratio, deviation of aspect ratio and temporal characteristics of similarity in histogram space using different measures (Bhattacharyya distance etc.) Experiments showed that especially a combination of aspect ratio and histogram features performs quite well. Due to the underlying tracking mechanism and the known correspondence between tracking and activity recognition, both the activity and its location where it happened is transmitted for further visualization and fast response (e.g., in case of falling persons). No video is necessary to be transmitted.

6.6 The Classification & Activity Recognition Flow

The whole flow for using the SVM engine on the camera is split in two phases:

- **1. Training phase** Within the training phase the supervised learning process is carried out. Here, an optimized kernel is generated and trained.
- **2. Online phase** The online phase applies during use of the system.

These phases are also differentiated in terms of speed requirements. To ensure maximal flexibility an approach is taken where the smart camera behaves like "hardware-in-the-loop". The training phase is transparently outsourced to the server to fully exploit the latest SVM optimization frameworks available today without the need of porting them to the smart camera. All that is required on the smart camera are efficient implementations of the feature extractors and the SVM online evaluation. The former is not necessary to be reimplemented on the server, instead the feature extraction runs transparently on the camera also within the training phase. This ensures that exactly the same extractors are used online as within training and avoids double implementations with the corresponding danger of bugs. From the camera perspective, hardware-in-the-loop here can thus be interpreted as "server-in-the-loop" compared to a full smart camera implementation. After the training phase the kernel is downloaded to the embedded SVM engine on the smart camera for fully embedded online evaluation.

6.6.1 Training Phase

The training is performed using the SPIDER framework from MPI, Tübingen [45] which is an object oriented SVM framework in MATLAB. After choosing a kernel function (e.g., RBF) the kernel parameters (σ, C) are optimized using a grid search with exponential steps in conjunction with cross validation. To this end, the performance under different combinations of kernel parameters is evaluated and the one with the best cross selection precision is chosen and trained. Cross validation is a preferred method especially if only few training examples are available. The loss of the kernel considered as best is computed. Within this framework the classification performance of different feature combinations can easily be tested and compared. Finally, the kernel together with the optimized parameters are downloaded to the camera in the field.

6.6.2 Online Phase

Within the online phase the whole SVM classification process is performed embedded on the camera. We chose to use SVMs not only due to their theoretical and practical benefits in terms of classification performance but also due to their relatively moderate computational demands in the online phase: During runtime, basically only a dot product between the feature vector combination of the actual camera image and each support vector has to be computed for each iteration. If the feature extractors are discriminative for the specific application and the training examples are representative, only few support vectors are necessary, thus speeding up the classification process. Examples both in terms of classification and activity recognition are depicted in the following.

6.6.3 Activity Recognition at The Assisted Living Home

Fig. 10 shows a sequence of a person tracked and analyzed within the hallways of the assisted living demonstrator. The event of falling is detected which is finally visualized as stylized icon as shown later in the web visualization (Fig. 17) and in the Google Earth visualization (Fig. 15). The falling person's position is marked with a

Fig. 10 Detected falling sequence at the assisted living home.

red warning icon and transmitted as a text message to a given phone via the alarm handler. The corresponding extracted feature vectors from this video sequence of the relevant camera are depicted in Fig. 11 where the detected falling event is marked. Encouraging results are made with the combination of aspect ratio, aspect ratio deviation and histogram differences. It showed that the best matching kernel is a RBF-Kernel. The loss after a three-fold cross validation was 0.64%. The training data was extracted from 11 different video files, wherein the amount of frames labeled as falling, compared to those labeled as not falling, is unbalanced. Our resulting kernel consists of 37 support vectors of 1508 training vectors (2.45%) which indicates

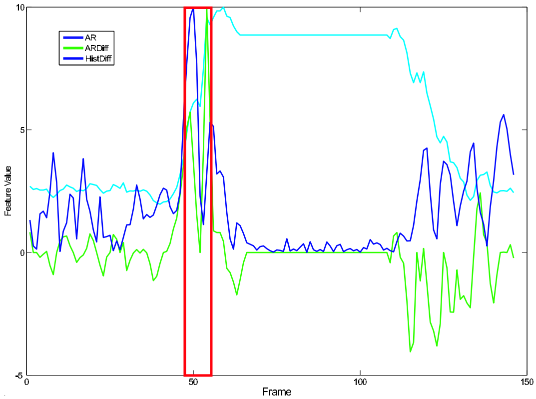

Fig. 11 Features over time, falling sequence is marked.

good generalization performance and discriminative features. Moreover, the same kernel can be used for all cameras installed in the assisted living home. The complete prototype system with activity recognition enabled runs on the smart camera node with 8-10 fps. Feature values over time are shown in Figure 11, the frames which indicate a falling person is marked red. Figure 10 shows the corresponding scene where the person is detected as falling. The system shows quite promising recognition performance, however the positive training examples (persons actually falling) are not enough to provide a quantitative performance analysis yet.

7 Visualization Fully Decoupled from Sensor Layer

We present three different visualization options which are designed for a more intuitive understanding of the surrounding, decoupled from the sensor network.

7.1 3D Visualization Node - XGRT

This visualization option enables that results of the camera network are directly embedded within a 3D model. To this end, we developed a mobile model acquisition platform called “The Wägele”[1]. It allows for easy acquisition of indoor and outdoor scenes: 3D models are acquired just by moving the platform through the scene to be captured. Thereby, geometry is acquired continuously and color images are taken in regular intervals. Our platform (see Fig. 12) comprises an 8 MegaPixel omnidirectional camera (C1 in Fig. 12) in conjunction with three laser scanners (L1-L3 in Fig. 12) and an attitude heading sensor (A1 in Fig. 12). Two flows are implemented

[1] *Wägele* – Swabian for a little cart

to yield 3D models: a computer vision flow based on graph cut stereo and a laser scanner based modeling flow. After a recording session the collected data is assembled to create a consistent 3D model in an automated offline processing step. More details can be found in [2, 17]. The visualization node gets its data from the server

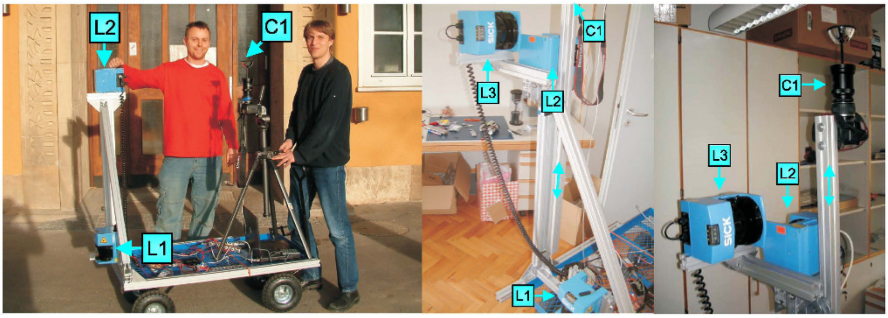

Fig. 12 The Wägele: Two setups of our mobile platform. **Left:** Two laser scanners L1, L2 and one omnidirectional camera C1. **Center & Right:** Three laser scanners L1, L2, L3 and omnidirectional camera closely mounted together.

node and renders the information (all objects currently tracked) embedded in the 3D model. This visualization option is based on the eXtensible GRaphics Toolkit (XGRT) developed by Wand et al. at our institute which is a modular framework for real time point based rendering [44]. Tracked objects can be displayed as sprites using live textures [18]. The viewpoint can be chosen arbitrarily, also a fly-by mode is available that moves the viewpoint with a tracked person/object. Results of the GRIS installation with both indoor and outdoor cameras is described now. First, a 3D model of both the indoor hallway and the outdoor environment has been acquired. Afterwards, the camera network has been set up and calibrated relative to each model. Fig. 13 illustrates results within the XGRT visualization. Each tracked persons (see Fig. 13,(1, 3)) is embedded within the acquired 3D model as live texture (Fig. 13,(2, 4)). Even under strong gusts where the trees were heavily moving, our per pixel noise process estimator enabled robust tracking by spatially adapting to the respective background movements.

7.2 Google Earth as Visualization Node

A geo referenced visualization of the scene analysis results embedded in Google Earth [21] is introduced in the following. The produced results of the surveillance system are permanently updated on a password protected web server which can be accessed ubiquitously by multiple users. According to the privacy needs of the application each embedded object is represented either as live texture or filtered

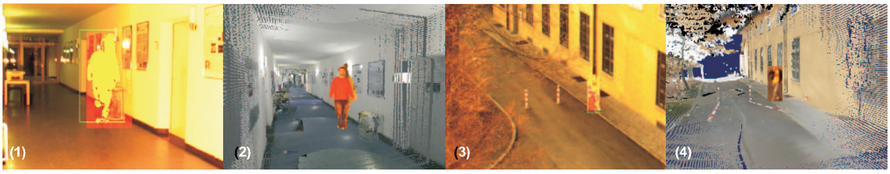

Fig. 13 XGRT visualization: **(1, 2):** Indoor experiment. **(3, 4):** Outdoor experiment. **(1, 3):** Live smart camera view with overlaid tracking information. **(2, 4):** Live XGRT rendering of tracked object embedded in 3D model acquired by the Wägele platform.

as icon within the Google Earth model, augmented by its status. The latter further increases the applicability of the system in areas with high privacy concerns, as the video data can stay completely in each smart camera. Tracking and activity results are updated at 20 Hz. Furthermore, in case of an recognized emergency a live overlay of the camera view can be displayed, updated at lower frequency (e.g. 1 Hz) to save bandwidth. Of course, the viewpoint within Google Earth can be chosen arbitrarily during runtime, independent of the objects being tracked (and independent of the camera nodes). A car tracking example is illustrated in Fig. 14: As a car approaches one camera's field of view, a new particle filter is automatically instantiated and tracks the car along its way. Within the Google Earth visualization the car is embedded as a stylized icon. Figure 15 shows the resulting Google Earth visualization of the home of assisted living, overlaid by a floor plan. Currently, one person is present within the perimeter covered by the prototype installation. It is detected as a moving object, automatically tracked and analyzed. In this moment, the person's activity is recognized as falling and thus marked by a red warning icon at the corresponding position in geo referenced world coordinates. No person specific data is visualized, the live view is just shown for demonstration purposes but could be enabled to appear in case of such a detected alarm situation. Not only the results of multiple cameras but even of multiple complete installations at different locations around the world could be visualized concurrently in this virtual world to provide a single-point of information. This addresses the problem of getting overstrained by the sheer mass of video data (*problem #2*), which is even more present if multiple installations have to be covered concurrently, is also addressed with this approach. At the same time, the visualization shows the benefit of the spatial information in world domain of the event, not just a result in 2D image domain ("falling person in camera #555"). As the viewpoint can be chosen arbitrarily during runtime, Fig. 16 shows a zooming sequence to a detected falling person in the assisted living home. This shall demonstrate how a scenario with multiple homes for assisted living could look like: Not only the results of multiple cameras but even of multiple complete installations at different locations around the world could be visualized concurrently in this virtual world to provide a single-point of information. This addresses the problem of getting overstrained by the sheer mass of video data (*problem #2*, which is even more present if multiple installations have to be covered concurrently, is also addressed with this approach. At the same time, the visualization shows the benefit

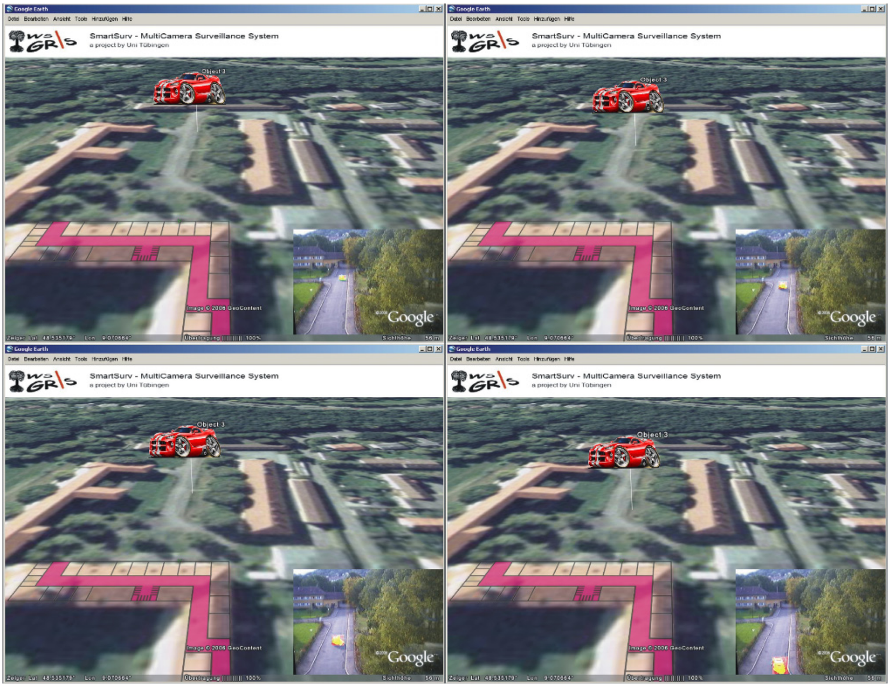

Fig. 14 Live tracking results embedded in Google Earth. The approaching car is tracked as "object3" in world domain, visualized as stylized car icon. An overlaid video feed of one camera node's output is shown on bottom right, respectively.

of the spatial information in world domain of the event, not just a result in 2D image domain ("falling person in camera #555").

7.3 Web Application for Visualization

An interactive web visualization application has been developed to cover users with just a web browser independent of Google Earth. Thereby, the system is ubiquitously available to any user, at the same time protected by standard secure techniques. Fig. 17) depicts the web visualization. In contrast to regular synchronous web page calls, only incremental changes are updated by rendering objects and activities in the floor plan on-the-fly. The update frequency can be chosen at runtime up to 10 Hz. No privacy critical data is shown unless an emergency is recognized. In this case the live view of the associated camera is automatically integrated at lower frequency (e.g., 1 Hz). To summarize, using the latter two ways of visualization, the resulting information from the distributed camera network is first integrated in one consistent world model which is then made accessible worldwide for multiple users independently.

Fig. 15 Google Earth visualization of the assisted living home prototype system. A person is tracked and classified as falling by the smart camera network – reflected within the Google Earth visualization located in world coordinates. On the right the live view of the respective smart camera is shown, overlayed by the 2D tracking output.

7.4 Statistics

The information within the virtual world model results in the following statistics. Fig. 18 illustrates the number of persons tracked concurrently within assisted living prototype over 24 hours of operation. In Fig. 19 the distribution of number of events per hour within a day is shown, accumulated over an entire week. As we do not have an automated way of detecting false tracking results, no qualitative results showing false object detection rate are available yet. One can see an early start in the day, a significant peak around lunch time and a very early drop at 7p.m. where the elderly persons already go to their rooms.

To further address problem #2 – the overstraining data problem – a timescaled visualization functionality is available. Thus, besides the live view of the current situation reflected in the virtual world model it is possible to access past time spans

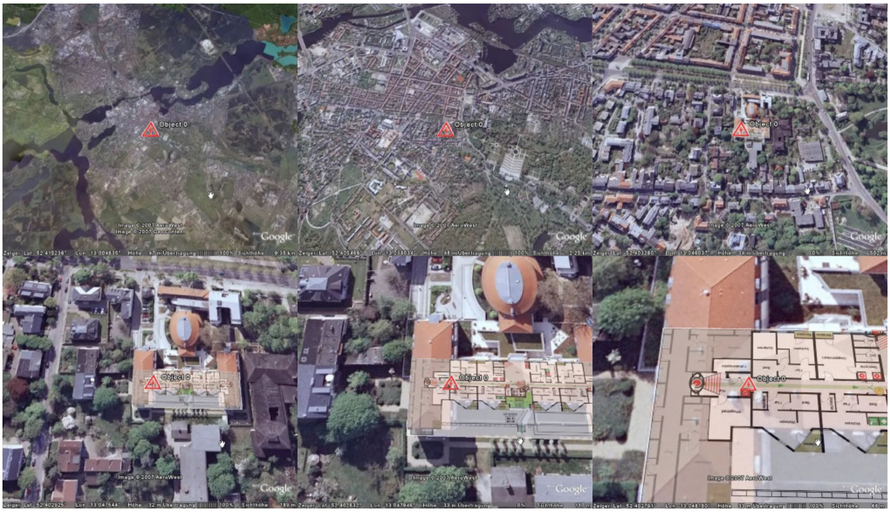

Fig. 16 Google Earth visualization: Zooming in to a person recognized as fallen.

Fig. 17 Ground plan with embedded camera information, tracking and activity recognition results. Every camera's pose in conjunction with its field of view is depicted within this plan. Concurrently, four persons are tracked by the smart camera network whereas one person has been falling (red icon on the right).

and get a quick overview of the situation, e.g., of the last 24 hours in the home for assisted living. The user can jump between events (tracking/activity recognition results), only time spans where events happened are logged for visualization. All events within the chosen interval width are shown concurrently to reduce the chance of missing events in case of fast forward. A temporal ("replay") animation in a definable speed is available.

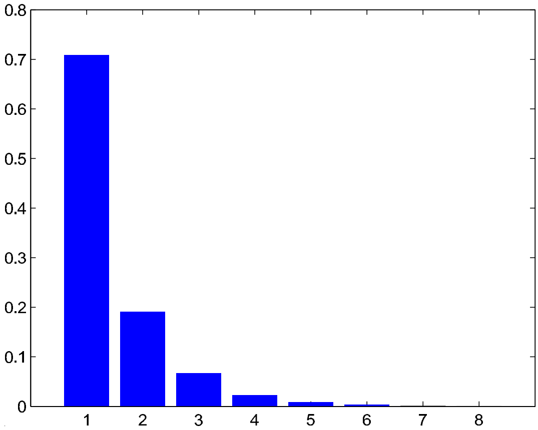

Fig. 18 Distribution of number of persons concurrently tracked by the camera network over one week of operation.

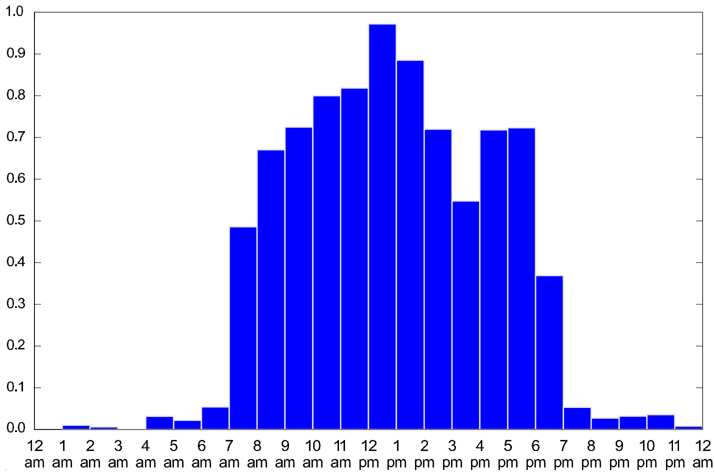

Fig. 19 Events over 24h within the whole perimeter covered by the camera network at the home for assisted living.

8 Conclusion and Impact

Surveillance applications are booming in various fields of applications. However, current manual and centralized automated systems impose several strong drawbacks, especially in terms of privacy. Three problems have been identified: lack of privacy, sheer mass of raw video data overstraining the operators, and limited flexibility of today's systems. A ***distributed*** and ***automated*** smart camera based approach has been proposed that addresses these problems. The approach is based on the smart camera paradigm where a camera behaves like an application specific sensor that just transmits the results which are on a higher level of abstraction. The presented architecture and system is capable of distributed tracking in geo referenced world coordinates. Furthermore, it features activity recognition and is extendible by arbitrary events as long as according features are designed. To this end, each smart camera is capable of embedded classification using Support Vector Machines (SVMs). Only results are then transmitted to ensure privacy which are embedded in one consistent geo referenced world model for visualization. A real world prototype system at a home for **assisted living** has been set up. It has been running 24/7 for several months now and shows quite promising performance. As a first activity, falling person detection has been addressed. An intuitive *visualization* system with three innovative options is presented where all information of the sensor systems is integrated in one consistent and geo referenced world model – the virtual world. The user is fully decoupled from the camera level and just sees the resulting movements within a geo referenced map. The idea is that he does not have to care about the underlying processes (which camera a person currently tracks and when a handover occurs), he just sees the path of each person in a world map, enriched by its status of the activity recognition.

8.1 Future Directions

Future work includes a systematic survey of the inhabitants of the assisted living home to gain more insight in how such a system is seen from the user's perspective. Moreover, it goes without saying that also challenges of the proposed approach have to be mentioned. The performance of the system relies on the performance of the automated video analysis algorithms. These do not complement the human operators but replace them from sensor level all the way up to a level where the information is not directly privacy related any more. The trust in the camera system from both sides is to be made clear: The persons being surveilled have to trust that it really does work as privacy respecting as advertised. With CCTV-style systems the user can at least be aware that he is currently being watched (now that he does not have a choice). At the same time, the operator has to trust in video analysis and self-diagnosis capabilities as no method of control is available to check its operativeness. One solution could be to certify the manufacturers of such systems as "trusted". Additionally, a self-diagnosis unit within the camera should check if it is still fully operational (e.g., by checking if the lens is still clean). The most important remaining challenge is, that automated video analysis is still a hot research topic with unsolved problems. Especially, the more abstract activities involving multiple persons and objects are topics of future research. Moreover, fully distributed and collaborative processing and decision making is a major research direction for future work. This could increase the robustness significantly, both on algorithmic level in terms of computer vision and with respect to fault-tolerance on hardware level. Performing the handover process in a distributed way is a first area to start with. Fusing object classification and activity recognition results is also an interesting future direction. Moreover, instead of using the discrete privacy filter options, i.e., either raw video texture or stylized icons, a continuous de-identification filtering could be used to get information in between by combining the presented architecture with techniques like the “respectful camera” and “de-identificated faces” mentioned in Section 3.1.2. According to the importance level of the activity (e.g. falling) and status of the user (e.g., a medical doctor) a more fine-grained access to privacy related data could be given in cases where video data should still be of interest.

Acknowledgements We would like to thank the students Florian Busch, Roland Loy, Florian Ohr, Christian Vollrath and Florian Walter for their contributions, Peter Biber for the joint work on the Wägele platform, Benjamin Huhle for many fruitful discussions and Matrix Vision for their generous support and successful cooperation.

References

[1] Aghajan, H.K., Augusto, J.C., Wu, C., McCullagh, P.J., Walkden, J.A.: Distributed vision-based accident management for assisted living. In: T. Okadome, T. Yamazaki, M. Makhtari (eds.) ICOST, *Lecture Notes in Computer Science*, vol. 4541, pp. 196–205. Springer (2007)

[2] Biber, P., Fleck, S., Wand, M., Staneker, D., Straßer, W.: First experiences with a mobile platform for flexible 3D model acquisition in indoor and outdoor environments – the Wägele. In: 3D-ARCH’2005: 3D Virtual Reconstruction and Visualization of Complex Architectures. Mestre-Venice, Italy (2005)

[3] Bramberger, M., Doblander, A., Maier, A., Rinner, B., Schwabach, H.: Distributed embedded smart cameras for surveillance applications. IEEE Computer **39** (2006)

[4] Chalimbaud, P., Berry, F.: Embedded active vision system based on an FPGA architecture. EURASIP J. Embedded Syst. **2007**(1), 26–26 (2007). DOI http://dx.doi.org/10.1155/2007/35010

[5] CNN: Smart cameras spot shady behavior. `http://edition.cnn.com/2007/TECH/science/03/26/fs.behaviorcameras/index.html` [Accessed 1.3.2008] (2008)

[6] Collins, R., Lipton, A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O.: A system for video surveillance and monitoring. Tech. Rep. CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2000)

[7] Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: Detecting objects, shadows and ghosts in video streams by exploiting color and motion information. In: ICIAP ’01: Proceedings of the 11th International Conference on Image Analysis and Processing, p. 360. IEEE Computer Society, Washington, DC, USA (2001)

[8] Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with hsvcolor information. In: Intel-

ligent Transportation Systems, 2001. Proceedings. 2001 IEEE, pp. 334–339 (2001)

[9] Cucchiara, R., Prati, A., Vezzani, R.: A system for automatic face obscuration for privacy purposes. Pattern Recognition Letters **27**(15), 1809–1815 (2006)

[10] Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. Expert Systems **24**(5), 334–345 (2007). DOI 10.1111/j.1468-0394.2007.00438.x. URL `http://dx.doi.org/10.1111/j.1468-0394.2007.00438.x`

[11] Datz, T.: What happens in Vegas stays on tape. `http://www.csoonline.com/read/090105/hiddencamera_vegas_3834.html` [Accessed 1.3.2008] (2008)

[12] Dietrich, D., Garn, H., Kebschull, U., Grimm, C., Ben-Ezra, M. (eds.): Embedded Vision Systems, vol. 2007. EURASIP Journal on Embedded Systems (2007)

[13] Dishman, E.: Inventing wellness systems for aging in place. Computer **37**(5), 34–41 (2004). DOI http://doi.ieeecomputersociety.org/10.1109/MC.2004.1297237

[14] Dufaux, F., Ebrahimi, T.: Scrambling for video surveillance with privacy. In: Conference on Computer Vision and Pattern Recognition Workshop on Privacy Research in Vision (CVPRW-PRIV), p. 160. IEEE Computer Society, Washington, DC, USA (2006). DOI http://dx.doi.org/10.1109/CVPRW.2006.184

[15] Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.S.: Background and foreground modeling using nonparametric kernel density for visual surveillance. In: Proceedings of the IEEE, vol. 90, pp. 1151–1163 (2002)

[16] of Engineering, T.R.A.: Dilemmas of privacy and surveillance - challenges of technological change. Tech. rep., The Royal Academy of Engineering, London (2007). URL `http://www.raeng.org.uk/policy/reports/pdf/dilemmas_of_privacy_and_surveillance_report.pdf`

[17] Fleck, S., Busch, F., Biber, P., Straßer, W.: Graph cut based panoramic 3D modeling and ground truth comparison with a mobile platform - the Wägele. In: IAPR Canadian Conference on Computer and Robot Vision (CRV 2006). Quebec City, Canada (2006)

[18] Fleck, S., Busch, F., Straßer, W.: Adaptive probabilistic tracking embedded in smart cameras for distributed surveillance in a 3D model. EURASIP Journal on Embedded Systems, Special Issue on Embedded Vision Systems **2007**(29858) (2007)

[19] Fleck, S., Loy, R., Vollrath, C., Walter, F., Straßer, W.: SmartClassySurv - a smart camera network for distributed tracking and activity recognition and its application to assisted living. In: ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC-07). Vienna, Austria (2007)

[20] Fleck, S., Straßer, W.: Adaptive probabilistic tracking embedded in a smart camera. In: IEEE Embedded Computer Vision Workshop (ECVW) in conjunction with IEEE CVPR 2005. San Diego, CA (2005)

[21] Google: Google Earth. http://earth.google.com/ [Accessed 1.3.2008] (2008)
[22] Gross, R., Sweeney, L., de la Torre, F., Baker, S.: Model-based face de-identification. In: CVPR Workshop on Privacy Research in Vision (CVPRW-PRIV), p. 161. IEEE Computer Society (2006)
[23] Hampapur, A., Brown, L.M., Connell, J., Lu, M., Merkl, H., Pankanti, S., Senior, A.W., fe Shu, C., , li Tian, Y.: Multi-scale tracking for smart video surveillance. IEEE Transactions on Signal Processing **22** (2005)
[24] Hauptmann, A.G., Gao, J., Yan, R., Qi, Y., Yang, J., Wactlar, H.D.: Automated analysis of nursing home observations. IEEE Pervasive Computing **3**(2), 15–21 (2004). DOI http://dx.doi.org/10.1109/MPRV.2004.1316813
[25] Hengstler, S., Prashanth, D., Fong, S., Aghajan, H.: Mesheye: a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance. In: IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks, pp. 360–369. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1236360.1236406
[26] Johannes Widmer, M.G.: Spiegel Online: Die schöne neue Welt der Überwachung. http://www.spiegel.de/flash/0,5532,15385,00.html [Accessed 1.3.2008] (2008)
[27] Kampel, M., Aguilera, J., Thirde, D., Borg, M., Fernandez, G., Wildenauer, H., Ferryman, J., Blauensteiner, P.: 3D Object Localisation and Evaluation from Video Streams. In: F. Lenzen, O. Scherzer, M. Vincze (eds.) 28th Workshop of the Austrian Association for Pattern Recognition (AAPR), Digital Imaging and Pattern Recognition, pp. 113–122. Obergurgl, Austria (2006)
[28] Keshavarz, A., Maleki-Tabar, A., Aghajan, H.: Distributed vision-based reasoning for smart home care. In: ACM SenSys Workshop on Distributed Smart Cameras (DSC) (2006)
[29] Kleihorst, R., Abbo, A., Schueler, B., Danilin, A.: Camera mote with a high-performance parallel processor for real-time frame-based video processing. In: ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC-07) (2007)
[30] Philips: Philips Lifeline. http://www.lifelinesys.com/ [Accessed 1.3.2008] (2008)
[31] Prati, A., Mikic, I., Cucchiara, R., Trivedi, M.: Analysis and detection of shadows in video streams: A comparative evaluation. In: IEEE CVPR Workshop on Empirical Evaluation Methods in Computer Vision, Kauai (2001)
[32] Quaritsch, M., Kreuzthaler, M., Rinner, B., Bischof, H., Strobl, B.: Autonomous multicamera tracking on embedded smart cameras. EURASIP Journal on Embedded Systems **2007**, Article ID 92,827, 10 pages (2007). Doi:10.1155/2007/92827
[33] Ribeiro, P.C., Santos-Victor, J.: Human activity recognition from video: modeling, feature selection and classification architecture. In: Workshop on Human Activity Recognition and Modelling, HAREM 2005 (2005)
[34] Sangho Park, M.M.T.: Video analysis of vehicles and persons for surveillance. Intelligent and Security Informatics: Techniques and Applications (2007)

[35] Schiff, J., Meingast, M., Mulligan, D.K., Sastry, S., Goldberg, K.: Respectful cameras: Detecting visual markers in real-time to address privacy concerns. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2007)

[36] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.L., Ekin, A., Connell, J., Shu, C.F., Lu, M.: Enabling video privacy through computer vision. IEEE Security and Privacy **3**(3), 50–57 (2005). DOI http://doi.ieeecomputersociety.org/10.1109/MSP.2005.65

[37] Shah, M., Javed, O., Shafique, K.: Automated visual surveillance in realistic scenarios. IEEE MultiMedia **14**(1), 30–39 (2007). DOI http://dx.doi.org/10.1109/MMUL.2007.3

[38] Siebel, N., Maybank, S.: The advisor visual surveillance system. In: ECCV 2004 workshop Applications of Computer Vision (ACV) (2004)

[39] Staff, S.E.: Integrator unveils security system for Wynn casino. `http://www.securityinfowatch.com/article/article.jsp?siteSection=344&id=9566` [Accessed 1.3.2008] (2008)

[40] Tabar, A.M., Keshavarz, A., Aghajan, H.: Smart home care network using sensor fusion and distributed vision-based reasoning. In: ACM International Workshop on Video Surveillance & Sensor Networks, VSSN 2006 (2006)

[41] Toereyin, B.U., Dedeoglu, Y., Cetin, A.E.: HMM based falling person detection using both audio and video. In: MUSCLE Network of Excellence Project (2006)

[42] Trivedi, M.M., Gandhi, T.L., Huang, K.S.: Distributed interactive video arrays for event capture and enhanced situational awareness. IEEE Intelligent Systems, Special Issue on Homeland Security (2005)

[43] Velalstin, S., Remagnino, P.: Intelligent Distributed Video Surveillance Systems. Institution of Engineering and Technology (2006)

[44] Wand, M., Berner, A., Bokeloh, M., Jenke, P., Fleck, A., Hoffmann, M., Maier, B., Staneker, D., Schilling, A., Seidel, H.P.: Processing and interactive editing of huge point clouds from 3D scanners. Computers & Graphics **2008**(1) (2008)

[45] Weston, J., Elisseeff, A., Bakir, G., Sinz, F.: Spider machine learning framework. `http://www.kyb.tuebingen.mpg.de/bs/people/spider/` [Accessed 1.3.2008] (2008)

[46] Williams, A., Xie, D., Ou, S., Grupen, R., Hanson, A., Riseman, E.: Distributed smart cameras for aging in place. In: Workshop on Distributed Smart Cameras, DSC 2006 (2006)

[47] Wolf, W., Ozer, B., Lv, T.: Smart cameras as embedded systems. Computer **35**(9), 48–53 (2002)

[48] Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. **38**(4) (2006). DOI 10.1145/1177352.1177355. URL `http://portal.acm.org/citation.cfm?id=1177352.1177355`