Crisis Management Using Multiple Camera Surveillance Systems

Leon Rothkrantz
Delft University of Technology
The Netherlands Defense Academy
l.j.m.rothkrantz@tudelft.nl

ABSTRACT

During recent disasters such as tsunami, flooding, hurricanes, nuclear disaster, earthquake people have to leave their living areas for their own safety. But it proves that some people are not informed about the evacuation, or are not willing or able to leave or don't know how to leave the hazardous areas. The topic of the paper is how to adapt current video surveillance systems along highway and streets to semi-automatic surveillance systems. When a suspicious event is detected a human operator in the control room has to be alerted to take appropriate actions. The architecture of the system and main modules are presented in the paper. Different algorithms to detect localize and track people are published by the authors elsewhere but are summarized in the current paper. The system has been tested in a real life environment and the test results are presented in the paper.

Keywords

Video surveillance, crisis management, tracking algorithm, license plate recognition.

INTRODUCTION

During recent disasters such as tsunami, flooding, hurricanes, nuclear disaster, earthquake people have to leave their living areas for their own safety. But it proves that some people are not informed about the evacuation, or are not willing or able to leave or don't know how to leave the hazardous areas. It is difficult to control huge areas for people left behind. But nowadays many highways and streets in a city are monitored by a network of video cameras for traffic management. Such an existing infrastructure could be used for video surveillance in case of a crisis. Usually such cameras are connected to a control room where operators are supposed to monitor the computer screens, displaying the movies of the video cameras as indicated in Figure1. But the traditional passive video surveillance proves to be ineffective or impossible as the number of cameras exceeds the capability of human operators to monitor them. At TUDelft there is a project running on automatic surveillance systems. The aim of this research project reported in this paper was to develop such an intelligent surveillance system to support operators in the control room. When a suspicious event is detected a human operator has to be alerted to take appropriate actions.

The current video surveillance infrastructure is mainly used for traffic management. Individual cars are detected and tracked to compute their speed. Assessment of the traffic flow is used to assess incidents, speed control and to reduce traffic jams. In our case we are focused on individual isolated cars. Cars can be identified by their license plate and can be tracked by multiple cameras. Analysis of the tracks may reveal if the car is leaving the area, or if the car has mechanical problems or if the driver lost his route. By identification of the car by its license plate can be checked if the driver belongs to the region. Parked cars on exceptional places can also belong to the target group. In case the driver or one of his passengers leave the car the recorded tracks can be analyzed. In all cases an alert has to be send to the control room and operators have to take appropriate actions.

To summarize, the research goals of this paper are:

-Is it possible to use the available surveillance system along highway and streets in a city of multiple video cameras to monitor a hazardous area in case of a crisis?

-Is it possible to develop intelligent visual surveillance to replace the traditional passive video surveillance that is proving ineffective as the number of cameras exceeds the capability of human operators to monitor them?

-Is it possible to localize and track cars and assess the possible behavior of the car driver?

The outline of the paper is as follows. In the next section we present an overview of relevant literature. Next we present a model/architecture of our system and present test results of several components. We end the paper with a conclusion and directions for future work.

Fig. 1. Security employee inspecting video screens   Surveillance camera

RELATED WORK

The field of video surveillance is very broad. Active research is going on in subjects like face recognition, 3D object modelling, multi-camera camera setups and human behaviour analysis.
Hampapur et al. (Hampapur et al., 2005) give an overview of what aspects are important for large-scale video surveillance systems. The challenges that are pointed out are: combining multiple sources of information, automatic event detection and deploying systems with a large number of cameras (cost-wise).
Hameete et al. (Hameete et al., 2012) describe the surveillance system envisioned by the authors for the NLDA area in Den Helder, The Netherlands. The system uses video signals from the surveillance cameras to secure the area. It is a centralised approach. All recorded data from the cameras are transmitted to a central server in the control room for processing and interpretation.

The TREC Video Retrieval Evaluation (TRECVID-2009) is a conference series that started in 2001. The goal is "to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results". In their paper (Over et al., 2010) give an overview of the TREC Video Retrieval Evaluation (TRECVID-2009). The goal of this evaluation was "to promote progress in content-based exploitation of digital video via open, metrics-based evaluation". In 2009, sixty three research teams submitted a video recognition system. There were four tasks: high-level feature extraction, search (fully automatic, manually assisted, or interactive), copy detection and surveillance event detection. Each system had to be able to do at least one of these tasks. During the evaluation, the submitted systems are tested and the results are presented in this paper. Our attention was captured by the last task: surveillance event detection. The goal of this task is to recognize particular visual events of people. Ten types of events were specified. The used data consists of multiple synchronized camera views. The work of (Yokoi et al., 2009) presents the Toshiba system provided to the (TRECVID-2009). It explains how the following four components are implemented, namely change detection, human detection, human tracking and event detection.
Yang et al. (Yang et al., 2009) "explored several novel technologies to help detecting high-level concepts". The local low level features are extracted using Scale-invariant feature transform (SIFT). The other feature used is called Space-Time Interest Points (STIP), which computes locations and descriptors for space-time interest points in video. Besides the mentioned features which capture both global and local properties of the scene, two additional features are used for some of the recognition tasks: extraction of Region of Interest (ROI) and the facial features.

A survey on contemporary remote surveillance systems for public safety is presented by T. Raty (Raty, 2010). This paper reviews the historical development of video surveillance and describes the three generations of surveillance systems. The focus is on generic surveillance, which is applicable to public safety. Public safety and homeland security are substantial concerns for governments worldwide. Recent events like terrorist attacks have increased the demand for security. Therefore, there is a growing interest for surveillance applications.
Junejo et al. (Junejo et al., 2007) state that a single camera is not sufficient to monitor a large area and propose a practical framework for an automatically configurable network of non-overlapping cameras that provides sufficient monitoring capabilities.
Li et al. (Li et al., 2009) state that the aim of multi-target tracking is to infer the target trajectories from image observations in a video. This poses a significant challenge in crowded environments where there are frequent occlusions and multiple targets have a similar appearance and intersecting trajectories. Bandini and Sartori

(Bandini and Sartori, 2005) present the concept of a monitoring and control system (MCS) which aims at helping humans in decision making regarding problems which can occur in critical domains. This is done by gathering data of the monitored situation, detecting if abnormal events happen and acting on it, often by alarming the operator. N. Ihaddadene (Ihaddadene and Djeraba, 2008) describes work that has been performed for the MIAUCE project. This project "aims to investigate and develop techniques to analyze the multi-modal behaviour of users within the context of real applications" (Djeraba, 2011). A part of this project has made an attempt to automatically detect abnormal behaviour at the exits of airport escalators using video cameras, which is described in this paper. The aim is to automatically detect congestions at the escalator exits. Tracking many objects with many sensors has been researched by H. Pasula, S. Russell, M. Ostland and Y. Ritov (Pasula et al.,1999). They describe the possibility of tracking highway traffic using multiple cameras. The multi-camera aspect is interesting, because it is a challenge how to track one object using several non-overlapping cameras and it is an important aspect of our research.

Conclusions drawn from this literature survey show that the field of video surveillance is very broad, it has many aspects that have all kinds of difficulties of their own. According to Raty (Raty, 2010), researchers have begun to consider architectures for video surveillance systems. In this paper we propose an architecture for a multi-camera video surveillance system.

ARCHITECTURE

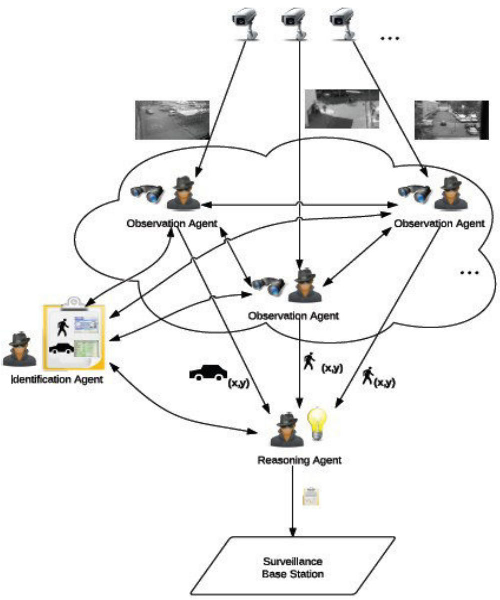

Fig. 2. Architecture of the system

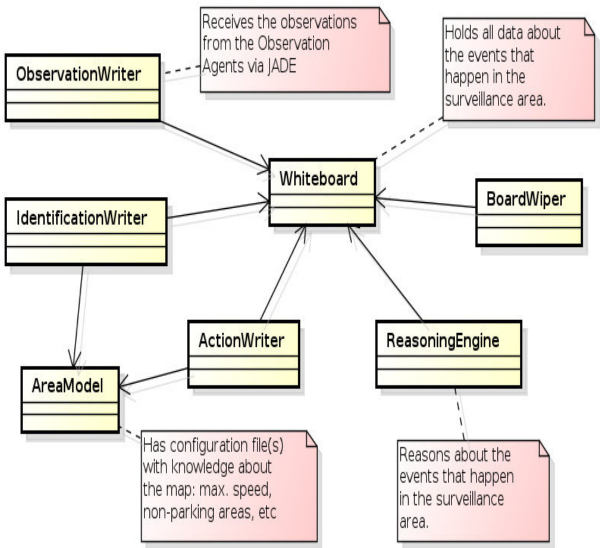

Fig. 3. Reasoning Agent summary class diagram

In Figure 2, 3 and 4 we depict our employed architecture. This architecture shows an agent network, composed of three kinds of agents. Each of them is described below:

The Observation Agent receives the video signal, processes it and sends object information to the Reasoning Agent. While processing it, it determines the track of the object and the class of the object (Car/Person). Furthermore, it asks the Identification Agent for a unique identifier (ID) of this object. Each Observation Agent knows which other Observation Agent processes a video stream that is physically close. To take some load of the Identification Agent, object descriptions are sent to the closest Observation Agents, since the just-discovered objects are likely to be observed by the neighbour Observation Agents too.

The Identification Agent handles requests from the Observation Agents. Internally, this agent has a list of all objects it encountered. For each object, characteristics are known such as class, colour, shape, etc. These characteristics are derived from the information received from the Observation Agents. For example, an Observation Agent could send a single image of an object to the Identification Agent. Then the Identification Agent matches it with the known objects to see if it is a new object or not. Finally, it sends a unique identifier to the Observation Agent which can be used in the system (for example to describe the object to the Reasoning

Agent). The advantage of such an agent is that there is only one entity that is responsible for matching objects. Knowledge about the map is asked from the Reasoning Agent.

The reasoning Agent receives the object descriptions and trajectories from the Observation Agent and reasons about this information. Mainly, it checks if the object has shown suspicious behavior using a few predefined suspicious events. Depending on the repetition and seriousness of these events, an alert is sent to the Surveillance Base Station. In order to provide more details of the Reasoning Agent, the software design is described here. The complete Reasoning Agent can be summarized by the UML diagram of Figure 3. The function of the Receiver is assigned to the ObservationWriter class. The observations are written to the Whiteboard, which corresponds to the Event Database. The roles of the Interpreters are filled by the IdentificationWriter and the ActionWriter. They add more information to the Whiteboard. The ReasoningEngine, which encompasses the Reasoning part, reads the contents of the Whiteboard and then reasons about it. Conclusions (alerts) are sent to the Alerting Agent by the ReasoningEngine. The Sender is included in the ReasoningEngine for simplicity. The BoardWiper is introduced to limit the memory that Whiteboard occupies. The Whiteboard is required to have 30 minutes of data and the BoardWiper regularly cleans the old data.

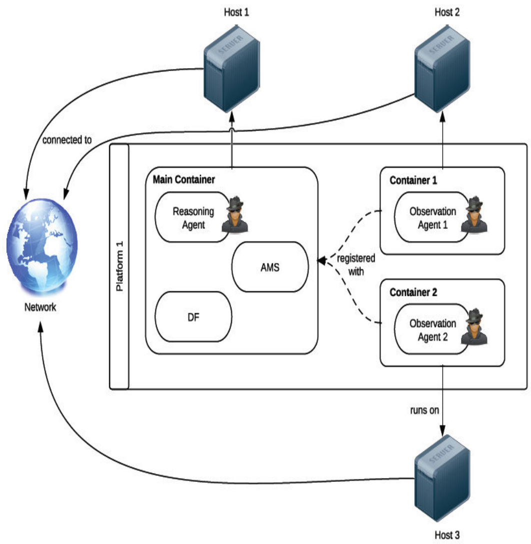

Fig.4. Structure of the Jade network

To implement our agents we have chosen the JADE architecture. Figure 4 illustrates how the JADE agents are organized. A JADE network can consist of one or more platforms. Each platform contains a Main Container and zero or more regular containers. A container can contain one or multiple agents. The Main Container always contains the Agent Management System (AMS) and the Directory Facilitator (DF). In short, the AMS provides a naming service and is the authority of the platform while the DF helps agents find each other. Each regular container within a platform is registered with the Main Container. The agent network in the example figure runs on three hosts. Host 1 contains the Main Container and the other two hosts have a container with an Observation Agent. The agents communicate via the network. New Observation Agents can be added to the regular containers if they run on the same host. If a new host is introduced, it creates a new container for his Observation Agent(s). The Alerting Agent is not shown in the figure. As mentioned above, JADE networks can contain multiple platforms. Agents can communicate with each other regardless of the platform they are a member of. The main advantage of this feature is that a JADE platform can be connected to non-JADE platforms as long as they obey the FIPA specification.

IMPLEMENTED MODULES

The main goals of our crisis surveillance system are to localise, identify and track moving objects as cars and pedestrians. Many tools have been developed by many researchers. We have implemented and tested several modules which will be described in the next section:

Localization and identification of cars

In (Cornet and Rothkrantz, 2003) we published our car license plate recognition system. A model for an image vision system is presented that is capable of recognizing car license plates independent from plate location, size, dimension, colour and character style. We proposed the application of a combined Neocognitron type of neural network classifier in generic car license plate recognition (CLPR) system. The suggested system contains an image-processor, a segment-processor and five combined Neocognitron network classifiers which act as a character recognizer.

The presented model of the system depends neither on specific license plate image features nor on license plate character style and size. Combining Neocognitron classifiers was motivated by the fact that manually tuning a training-set for a large Neocognitron network is tedious. It is shown how training set tuning for a large Neocognitron network can be avoided. By connecting small Neocognitrons specifically trained for character classes that are frequently wrong classified, the performance of the recognizer in our CLPR was improved significantly. The use of a Neocognitron recognizer contributes significantly to the generality of a CLPR system. Besides, character recognition achieves over 98% accuracy, using the proposed Neocognitron configuration.

Car tracking

Our first car tracking module (Lefter et al., 2010) is based on the mean shift algorithm implemented in OpenCV (Comaniciu 200). It consists of 5 processing components. The foreground/background discriminator (FG/BG) labels each pixel as either foreground or background. The next module in the pipeline is the Blob Entering Detection module. It makes use of the result (FG/BG mask) of FG/BG Detection in order to detect a new blob object that has entered a scene on each frame. At this stage the Blob Tracking module is initialized and it results in tracks of each new entered blob. Figure 5 illustrates a picture of a tracked moving car.

Fig. 5. Tracked car

As second tracking module, we adapted a tool called Predator (Kalal et al., 2010) also known as OpenTLD. The software tool Predator uses a new technique called P-N learning. The main idea is that the tracker learns appearances of the object (positive examples) and its direct neighbourhood (negative examples). The algorithm is able to learn from its mistakes. The task of the P-N learning algorithm is to learn a classifier that labels each unlabeled sample from a feature set using an a priori labeled set. This task comes down to estimating the parameters which are learned from a training set, similar to supervised learning. However, unlike supervised learning, P-N learning iteratively augments the training set by examples from the constraints of unlabeled data. During the training procedure, each iteration assigns labels to unlabeled examples using the classifiers trained in the previous iteration. "The constraints are then used to verify if the labels assigned by the classifier are in line with the assumptions made about the data. The example labels that violate the constraints are corrected and added to the training set. The iteration is finished by retraining the classifier with the updated training set. This procedure iterates until convergence or other stopping criterion". The conclusion is that the imperfect positive and negative examples cancel their errors under certain conditions. This algorithm is applied to video object detection. The problem is stated as follows: "given a single example of an object, learn an object detector on-line from unlabeled video sequence".

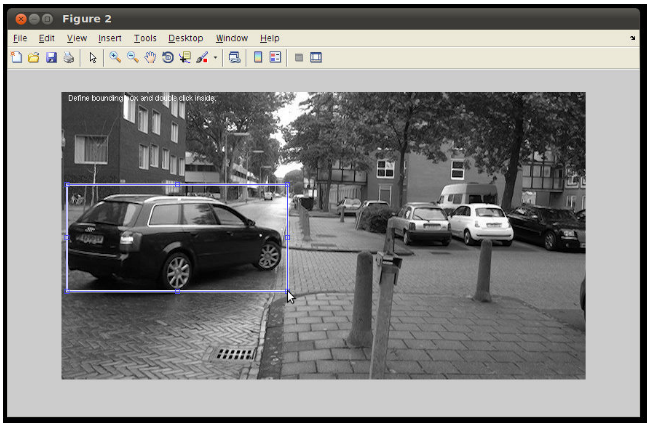

Fig. 6. Object localisation

To localise cars as indicated in Figure 6 an object detector is used which uses the scanning window strategy. A Lucas-Kanade tracker (Lucas et al., 1981) is used to track the object from frame to frame. This tracker is evaluated using the confidence determined by a normalized cross-correlation between the tracked patch and the patch selected in the first frame. Processing the video sequence is done as follows: for each frame, both the detector and the tracker find the location(s) of the object. “The patches close to the trajectory given by the tracker and detections far away from this trajectory are used as positive and negative examples, respectively” (Kalal et al. 2010). If the trajectory is considered valid (the tracker has 80% confidence in the last frame), these examples are used to update the detector. However, if the trajectory is invalid, the examples are discarded and the tracker is re-initialized. The algorithm is applied to synthetic data and on real data. The former experiment shows that using both P and N examples make that errors cancel out, whereas using either P or N give worse results. The experiment on real data is performed on six video sequences that are all used in experiments of five other papers. For five out of six video sequences, the algorithm manages to keep track of the object longer than algorithms described in all other papers.

For human classification a Support Vector Machine (SVM) was used in conjunction with an algorithm called Histogram Oriented Gradients (HOG) (Popa et al. 2012). This algorithm computes a feature vector of the given input image, which is then used by SVM for classification. The Support Vector machine was trained on the MIT pedestrian dataset1. The HOG algorithm computes the oriented gradient of the input image and overlays the image with a grid of cells. Then, for each cell a histogram is created with bins based on the orientation, where each bin is weighted by the magnitude of the gradient. Next, overlapping blocks are created consisting of adjacent cells and the cells in each block are contrast normalized to get rid of the change in light intensity over the image. The histograms are then put into a single vector, which is the feature vector used for classification. For a more comprehensive explanation of the histogram oriented gradient algorithm we refer to the original paper by (Dalal et al. 2005).

EXPERIMENTAL RESULTS

We tested the different modules of our surveillance system and we report next the obtained test-results for each individual module.

Fig. 7. Example of recordings

Car tracking

We made 100 video cameras on different days, time of the day of passing cars as indicated in Figure 7.The video camera was attached to a lamppost. We used the mean-shift algorithm to detect if a moving object is recorded at the borders of view window. This object is then selected and tracked by the Predator algorithm. In all of our experiments we consider new objects. Moving objects that were part of the background are avoided in the datasets. Furthermore, the Object Detector is expected to focus on the largest objects if multiple objects appear in the frame. Finally, given the implementation, we expect the Object Detector to detect objects that are at least two pixels away from the boundary of the frame. The performance was very good. Even in the case a second car appears it was tracked correctly as a second object. We notice that new objects always appear in the left lower corner or the right upper corner. This was learned by the Predator software. We noticed that our recordings are under good light conditions. The only problematic situation in our recordings was the case of two cars driving too close to each other. In those cases the two objects fused together to one object.

Pedestrian tracking

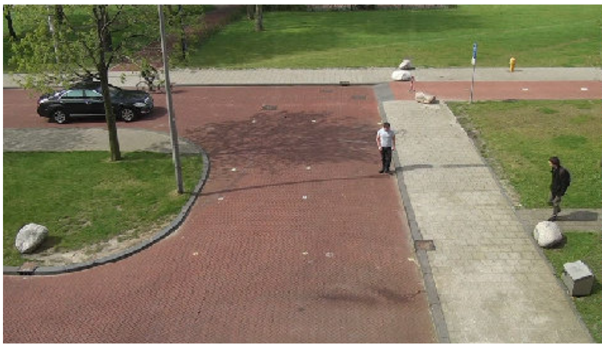

Fig. 8. Example of detected persons around a car

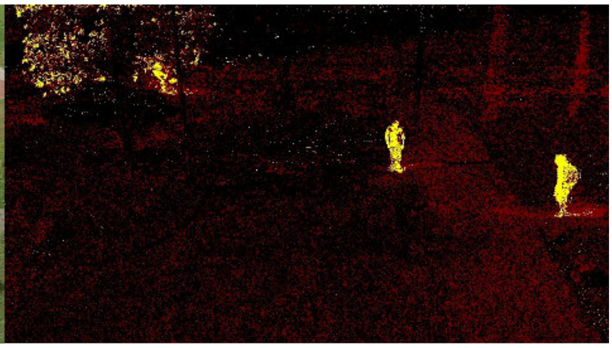

Fig. 9. A binary image resulting from the Mixture of Gaussians background subtraction

As expected pedestrians are small objects in the window (see Fig 8.9). They can enter the window from all sides. So the performance decreases considerably compared to cars. In case the Predator software was able to track the objects, we were able to analyze the tracks. Based on the recorded data we analyzed the tracks and the corresponding behavior:

- walking, running or stopping,
- walking in a straight line or taking a curved trajectory,
- changing directions (left/right, forward/backwards).

	Tracking	Walking, running stopping	Straight/curved trajectory	Left/right turns
Cars	100%	-	86%	92%
Pedestrians	82%	75%	84%	88%

Table 2. Results of tracking experiments

In Figure 10 we display the user interface for the human operators in the control room. The curved tracks are displayed and also some corresponding features as speed and curvature of the recorded tracks are depicted next to the map.

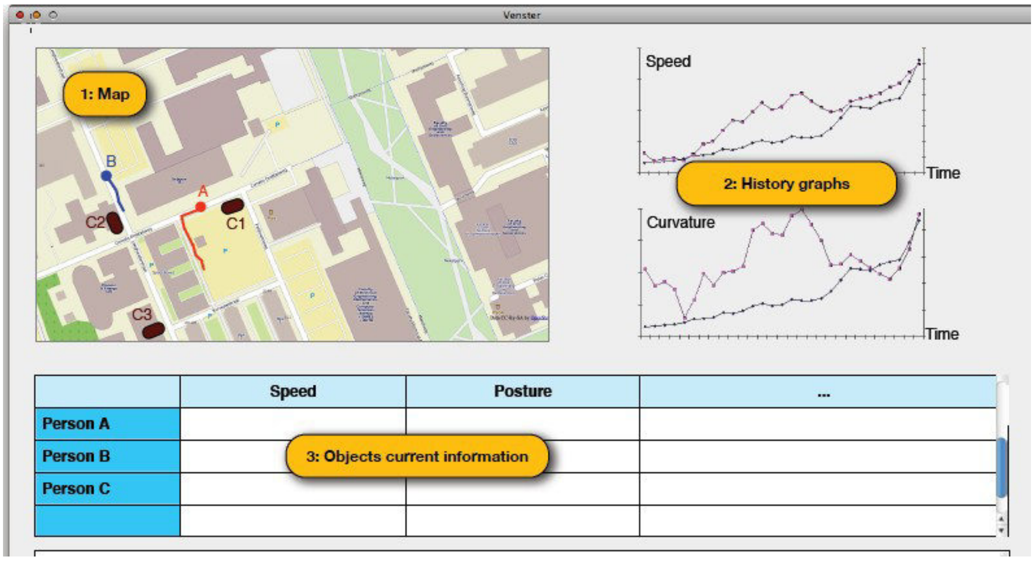

Fig. 10. Display of user interface in the control room

FIELD TEST OF THE SYSTEM DURING FLOODING

It is well known that more than 30% of The Netherlands is situated below sea level. Much new land was created as follows. First part of the water was surrounded by dikes and channels and next using windmills the inner water was removed via the channels. The new land is below sea level. During heavy rains the capacity of the rivers is not sufficient. To create a water buffer holes in the dikes are created and water floods in a controlled area. But first all the people and animals have to be evacuated from to be flooded area. In the winter of 2012 water buffers were created in the Northern part of the Netherlands and an area of more than 200 square km was flooded. It proved that many people were not willing to leave their homes even for their own safety. They wait until the very last moment when the water is already raising. Next many people want to return to their homes after some time. A lot of people are needed to supervise the flooded area and the many entry roads. The flooded area is crossed by many dikes and the roads on the dikes are above the water level and supervised by many cameras as indicated in Figure 11. But to supervise thousands of cameras in the control room staffed by limited operators and equipped by limited screens is almost impossible. Many police cars were crossing the area searching for people left behind, even helicopters were used. Experts agreed that a more effective and efficient solution is to transform the existing infrastructure of surveillance cameras in a network of automated surveillance cameras. If a car, cyclist is detected on the roads on the dikes an alert is send to the control room. Given the alarming situation the object will be tracked or a police car positioned in the area is informed to provide rescue service.

Researchers of TUDelft were requested to design such a surveillance system of smart cameras. The video recordings were used to train and test the system. It is a challenging task to detect, track moving objects on the roads under bad lighting conditions, heavy rains etc. But fortunately after mitigation the roads are not crowded anymore. Only isolated objects try to escape form the hazardous area. Analyzing the recorded video data it proves that isolated moving objects could be detected and tracked. License plates of cars could only be recognized if the cars were close to the cameras. The same holds for the detection of pedestrians and bicycles. Identification of objects is needed to track objects using multiple cameras.

The developed system will be used as a decision support system. The system generates alerts and these alerts trigger the human operator to check visually what is going on. False positives will be eliminated in this way. Operators in the control room stress the fact that they will be in charge and not overruled by a fully automated system. But they agreed that a surveillance system would be useful and even necessary. The surveillance task is a boring job in case only rare events happen. Maybe in the future the system can be fully automated. Another option to be developed is the wireless communication between the police cars and the smart cameras. An interesting option of the current system is that rescue workers in the hazardous area are also supervised by the cameras. In case of an incident this will be detected and a human operator in the control room has to decide if help is needed.

To conclude at this moment the system has been tested offline on video recordings of a controlled flooded area. Once the system has been installed it can be used in disaster situations when dikes break down on unexpected places and time. The camera network has been installed already, only minor technical add-ons are needed. It proves that the cameras situated on dikes are damaged by the raising water. The main advantage of the system, reported by the employees in the control room is increased situation awareness after the onset of a disaster caused by flooding.

Fig. 11. Flooded area (ANP-Press)

CONCLUSIONS

In the introduction we defined our research goals. In this section we will evaluate if we were able to realize these goals.

-Is it possible to use the available surveillance system along highway and streets in a city of multiple video cameras to monitor a hazardous area in case of a crisis?

We used the existing surveillance system in the harbor of Den Helder in the Netherlands to test our system. After installation of the cameras, implementation of the software modules and gathering information from the context we were able to run our experiments. The system is designed for (de-)central, distributed processing, but we were only able to test the central mode. The system was tested during average conditions. As mentioned in the last section recordings from the flooding in 2012 were used to test the pattern recognition software. In the next future we hope to be able to test the system during a (simulated) crisis situation.

-Is it possible to develop intelligent visual surveillance to replace the traditional passive video surveillance that is proving ineffective as the number of cameras exceeds the capability of human operators to monitor them?

Currently the system will be used to support the operators in the control room. Many areas in the harbor are closed for non-authorized people. The system was able to detect intruders in those areas monitored by cameras and send an alert to the control room. But the monitored area was limited and also the amount of cameras. As a future direction, we plan to test the system in the city of Amsterdam with a huge and dense network of surveillance cameras.

-Is it possible to localize and track cars and assess the possible behavior of the car driver?

Monitoring cars and people by surveillance cameras is a complex process and is very context sensitive. Extreme weather conditions and the quality of the cameras will have a huge impact on the results. But the same holds for the human operators in the control room. We stress the fact that we planned to monitor isolated objects so no crowd or traffic during the rush hours. The multiple cameras allow to process redundant data. Therefore possible errors from one camera at some moment can be corrected by recordings from another camera or at other moments of time.

REFERENCES

1. Bandini, S. and Sartori, F. (2005) Improving the Effectiveness of Monitoring and Control Systems Exploiting Knowledge-Based Approaches. *Personal and Ubiquitous Computing*, 9(5):301– 311.
2. Cheng, C. and Fan, G. (2006) What Can We Learn from Biological Vision Studies for Human Motion Segmentation? *Proc. International Symposium on Visual Computing*, Lake Tahoe, Nevada, 2006.

3. Comaniciu, D., Ramesh, V. and Meer, P. (2000) Real-Time Tracking of Non Rigid Objects using Mean Shift, *IEEE Conf. on Computer Vision and Pattern Rec*., vol. 2, pp. 142-149.
4. Cornet, B. and Rothkrantz, L. (2003) Recognition of car license plates using a Neocognitron type of artificial neural network, *Neural Network World,* vol. 13, no. 2, pp. 115-132.
5. Dalal N. and Triggs. B. (2005) Histograms of oriented gradients for human detection. In CVPR, pages 886--893.
6. Djeraba, C. (2011) MIAUCE Home Page. http://www.miauce.org/. Date accessed: April 14, 25.
7. Green, M. W. (1999) The appropriate and effective use of security technologies in U.S. schools: a guide for schools and law enforcement agencies, National Institute of Justice, 12, 13.
8. Haan de, G., Scheuer, J., Vries de, R., Post, F. H. (2009) Egocentric navigation for video surveillance in 3D Virtual Environments, *IEEE Symposium on 3D User Interfaces*, 12, 103–110.
9. Hampapur, S., Brown, A., Connell, L., Ekin, A., Haas, N., Lu, M., Merkl, H. and Pankanti (2005) Smart Video Surveillance: Exploring the Concept of Multiscale Spatiotemporal Tracking, *IEEE Signal Processing Magazine*, 22 2 38–51.
10. Hameete, P., Leysen, S., van der Laan, T., Lefter, I. and Rothkrantz, L. (2012) Intelligent Multicamera video surveillance, *International Journal on Information Technology & Security*, 4:51-62.
11. Ihaddadene, N. and Djeraba, C. (2008) Real-time crowd motion analysis. *19th International Conference on Pattern Recognition*, 25 1–4.
12. Junejo, I. N., Cao, X. and Foroosh, H. (2007) Autoconfiguration of a Dynamic Nonoverlapping Camera Network. *IEEE Transactions on Systems, Man, and Cybernetics -Part B: Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society*, 37(4), 21 803–816.
13. Kalal, Z. (2011) TLD, http://info.ee.surrey.ac.uk/Personal/ Z.Kalal/tld.html. Date accessed.
14. Kalal, Z., Matas, J., Mikolajczyk, K. (2010) P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 45-46-86-101 49-56.
15. Lefter, I., Rothkrantz, L., Burghouts, G. and Buchner, P. (2010) Multimodal car driver surveillance system in a military area. *Driver Car Interaction & Interface 2010*, Prague.
16. Li, Y., Huang, C. and Nevatia, R.(2009) Learning to Associate: HybridBoosted Multi-Target Tracker for Crowded Scene. *IEEE Conference on Computer Vision and Pattern Recognition*, 22 2953- 2960.
17. Lucas, B.D. and Kanade, T. (1981) An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop*, pages 121—130.
18. Over, P., Awad, G. and Fiscus, J.: TRECVID 2009 (2010)- Goals, Tasks, Data, Evaluation Mechanisms and Metrics, *Proceedings of TRECVID 2009*. NIST, USA, 16.
19. Popa, M., Rothkrantz, L, Shan, C., Gritti, T. and Wiggers, P. (2012). Semantic Assessment of Shopping Behavior Using Trajectories, Shopping Related Actions, and Context Information, *Pattern Recognition Letters*, DOI: 10.1016/j.patrec.20122012.04.015.
20. Popa, M., Rothkrantz, L., Wiggers, P. and Shan, C. (2012) Shopping Behavior Recognition using a Language Modeling Analogy*, Pattern Recognition Letters*, DOI:10.1016/j.patrec.2012.11.015.
21. Raty, T. D. (2010) Survey on Contemporary Remote Surveillance Systems for Public Safety, *IEEE Transactions on Systems, Man, and Cybernetics -Part C: Applications and Reviews*, 40(5), 20, 26 (2010) 493–515.
22. Valera, M. and Velastin, S. A. (2005) Intelligent distributed surveillance systems: a review, Image (Rochester, N.Y.), 20041147, 13.
23. Yang, X., Xu, Y., Zhang, R., Chen, E., Yan, Q., Xiao, B., Yu, Z., Li, N., Huang, Z., Zhang, C., Chen, X., Liu, A., Chu, Z., Guo, K., Huang, J. and Shanghai Jiao (2009) Tong University Participation in High-Level Feature Extraction and Surveillance Event Detection at TRECVID 2009, 17.
24. Yokoi, K., Watanabe, T. and Ito, S.: Toshiba at TRECVID 2009: Surveillance Event Detection Task.