APFEL - FAST MULTI CAMERA PEOPLE TRACKING

AT AIRPORTS, BASED ON DECENTRALIZED VIDEO INDEXING

Sebastian Hommel, Matthias A. Grimm,
Darius Malysiak, Uwe Handmann

Abstract:
The research project APFel deals with the development of an assistance system for security personnel. This system supports the human operator in taking security-critical decisions. It enables the search of possible whereabouts of a suspect in the past, the present and in the near future in nearly real-time. Biometric and clothing based features combined with statistical knowledge are used in consideration of a meaningful, spatio-temporal analysis to resort the recorded video data. This allows a quick assessment of a potentially dangerous situation. Costly and expensive false alarms can be avoided and safety-critical situations are resolved quickly.

Kurzfassung:
Das Projekt APFel beschäftigt sich mit der Entwicklung eines Assistenzsystems für Sicherheitspersonal. Dieses Assistenzsystem unterstützt bei sicherheitskritischen Entscheidungen. Es erlaubt die schnelle Suche nach möglichen Aufenthaltsorten einer verdächtigen Person in der Vergangenheit, Gegenwart als auch in der nahen Zukunft. Dabei werden biometrische und kleidungsbasierte Merkmale in Kombination mit statistischem Wissen und unter Berücksichtigung einer sinnvollen, räumlich-zeitlichen Betrachtung verwendet. Die hierdurch neu sortierte Videodatenmenge ermöglicht so eine schnelle Einschätzung einer potenziellen Gefahrensituation. Aufwendige und teure Fehlalarme können so vermieden und sicherheitskritische Situationen schnell geklärt werden.

Background

Ten years after the 9/11 terrorist attack the aviation security infrastructure and politics have changed drastically around the world. Nevertheless the threat situation remains unforeseeable and unpredictable and is tightened further by various attack methods of understaffed security personal. The maintenance of the security in these infrastructures, which besides the large international airport hubs also include numerous regional and General Aviation (GA) airports, is of the highest priority. The risk exposure of aircraft for terrorist attack methods as well as the occurrence of organized crime in the form of drug smuggling and human trafficking at airports is a problem that needs to be addressed. Large commercial airports possess a dedicated GA terminal and related facilities whose access to the security critical and tightly regulated infrastructure at the airport is often not sufficiently protected and kept under surveillance. In conclusion, adequate technological development is needed to account for the increased security requirements. The realization of these security regulations also need to be performed in a cost-efficient manner. Airports prefer to operate the protection of infrastructure at low costs and in an efficient manner by using a specially designed technology, in order to prevent the introduction of new and extended security regulations.

As one significant element in the security plan of an airport the video surveillance has been established. The size and complexity of modern airport infrastructures require an increasing number of cameras, which are digitally connected and provide images in high resolution. The resulting challenge is an efficient analysis of the video material within a short time span. Cases like those at the airports of Munich and Newark/New Jersey in January of 2010 have shown that besides modern video technologies, a re-identification of a once seen person over multiple camera viewpoints turns out to be a time-consuming and tedious task, which may even

lead to an expensive temporary shut-down of a terminal.

Objectives and Requirements

The overall objective of the project is the development of a system for the support of security personnel in video surveillance centers and security control stands. The essential function of the system is the video-based people recognition, which supports the human operator in analyzing recorded video data from several cameras (backward analysis), as well as to search a person at the life video streams (forward analysis) to perform a walk path reconstruction of a person in a non-overlapping camera network. Based on the experience of a security operator, the system is triggered by marking a person in a video frame (monitoring). With the aid of the system's video analytics it can be determined, which walk-path the person has already taken in addition to where the person is now. Thus helps the security personnel to estimate the threat potential. Two major requirements of the system are the efficient processing and thinning of different video data sets as well as the very high speed in analyzing image material, in order to reach a significant time advantage as opposed to the traditional way of a "manual" search. Furthermore, the system combines different video analytical subsystems for the identification and re-identification of persons to achieve robust results under different conditions.

Besides the technical aspects, the research project contains several socio-scientific aspects and the examination of requirements for a data privacy compliant layout. Among them are an analysis of the acceptance of passengers and its implications on their subjective security sense. Additionally, the usability and graphical interface of the system application is evaluated from the operators' standpoint.

In this project, concepts are prepared and tested to recognize which way a person is most likely to take. This is based on location probabilities by comparing typical patterns added with knowledge about the geometry and the timetable of an airport. The typical patterns are static recorded by a temporal installed laser network.[1]

System Design

To re-identify a person over multiple cameras, an intelligent video analysis is necessary. In order to process the resulting large amount of image specific data, fast and effective analytics logic with corresponding image preprocessing is integrated into each camera in a decentralized way. This enables an early and effective thinning of relevant image data. For the application scenario an additional higher level, cross-camera information is needed, which is realized by using central components. The exchange of metadata, which are extracted by subtasks of the system, such as the person detection, person tracking, feature extraction, face detection and tracking, person re-identification or prediction of person specific routes, is carried out via a central database and a message server.

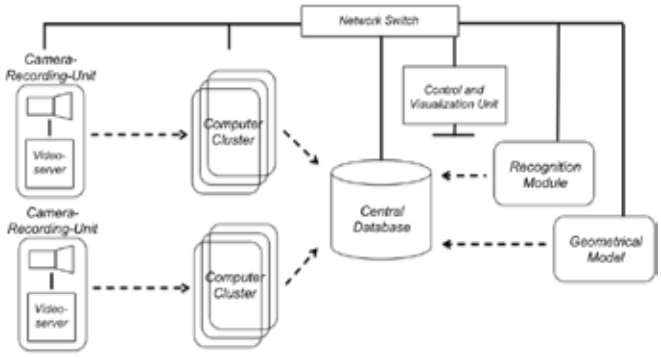

Fig. 1: Overview of the hardware architecture.

Decentralized components

In order to perform the data analysis in real-time, fast preprocessing steps are necessary to reduce the huge amount of image data. This preprocessing allows a later recognition of people and the prediction of the current and future location of a person. On a first abstraction level an attention-based detection of people and corresponding faces is performed. Therefore, adaptive foreground segmentation is combined with detection methods in such that people and faces are searched only in foreground areas. This speeds up computation time and reduces false detections. Furthermore, some first person features are extracted like color and texture of the clothing which supports the later backward analysis by reducing hypotheses. In addition to the preprocessing, a fast camera specific illumination correction is performed.

Illumination Correction

Illumination correction is the first decentralized component which is integrated before the images are recorded. This step is necessary to obtain features which are invariant to different cameras, daytimes and weather conditions (sunny, rainy, drifting clouds) in outdoor areas as well as changing of light indoors due to windows. This enables a consistently good performance even in dynamically changing conditions.

1. Kolarow, A., Brauckmann, M., Eisenbach, M., Schenk, K., Einhorn, E., Debes, K., Gross, H.-M. , Vision-based hyper-real-time object tracker for human-robot interaction, IROS, 2012

Fig. 2: without any image correction

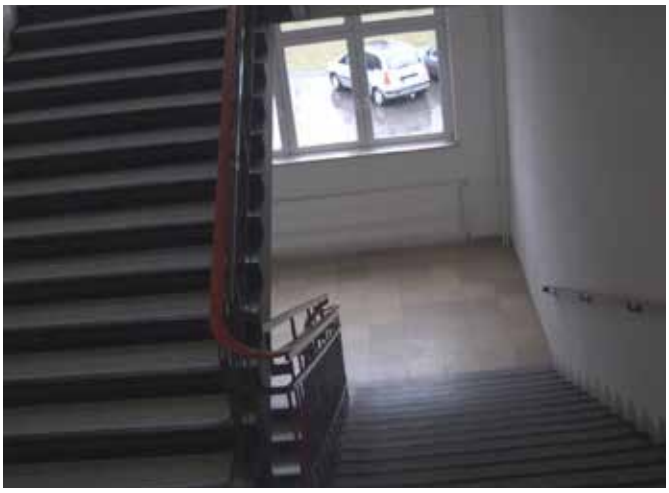

Fig. 3: with our illumination correction

Quelle: Hommel, S., Grimm, M. A., Voges, V., Handmann, U. and Weigmann, U., An intelligent system architecture for multi-camera human tracking at airports, in CINTI 2012

The used image enhancement is based on the camera internal pixel representation (36 bit) which is linearly mapped to the cameras output (24 bit). In our case, it is preferable to use an adaptive logarithmic mapping function to correct different illuminations (gamma correction[2]). By using this camera internal correction, the quantization noise does not increase since no sensor information is over represented. The value of gamma is estimated by the minimal pixel value of the gray-scaled image separately for each image. To handle short time illumination changes, it is preferable to smooth the gamma temporally. This gamma correction is used in combination with a low brightness threshold for the camera exposure time adaptation. In this way, the recorded images are darker with less overexposure in the first step, whereas the mapping function makes the image brighter and smoother in illumination. Thus the recorded 24bit image represents some of the previously overexposed image and black areas as well as the well illuminated areas. A further advantage of this method is the reduction of the motion blur by decreasing the exposure time.

Foreground-segmentation

After the illumination correction is performed, areas with people are indexed. For this camera related analysis, salient based people detection is used. The first and fastest step to reduce the data volume during the live analysis is the segmentation of foreground areas. One of the used methods is interconnected with an optimized, histogram oriented gradient based people detector (see next subchapter). This detector detects all people in the foreground areas before the model of the background is updated only in those areas, where no people are detected. This combination speeds up the detector and reduces false detections. Furthermore, this combination reduces the effect of learning no or less moving people into the background model. To handle differently illuminated scenes, a dynamic threshold will be calculated by using the standard variance of the gray scaled image. Furthermore, an opening on the binary foreground map is used to reduce the influence of noise. First the threshold to evaluate the difference between input image and background model is calculated and a binary foreground mask is determined. Now, the people detector works at the foreground areas and the background model is updated for all non-person areas.

People Detection

In terms of computation time, the detection process is one of the most time consuming steps within the tracking process. A reliably way of detecting people on an image is the method of histograms of oriented gradients[3]. This algorithms principle can be roughly summarized as follows. It extracts edge gradients from an image pixel by pixel and assigns each gradient into one of nine orientation bins for a small (e. g. 8px x 8px) image region. The orientation bins from each image region are then sequentially concatenated into a feature vector which in turn is being used as the input for a classifier trained for people detection. The output of the detection process is a set of regions of interest (ROI). Each ROI represents an image area which contains a person. The extraction of the feature vector spent the mentioned computation time for this method. In order to increase the reliability for a successful detection of humans, we applied multiple classifiers (e. g. head part, head-shoulder part etc.). Although we achieved a high rate

2. Scott, J., Pusateri, M., Towards real-time hardware gamma correction for dynamic contrast enhancement, in AIPR Workshop, pp. 1-5, 2009

3. Dalal, N., Triggs, B.,"Histograms of oriented gradients for human detection," in CVPR 2005, vol. 1, pp. 886–893.

Fig. 4

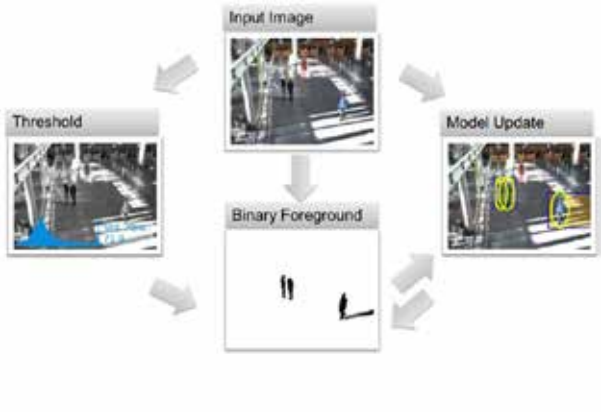

Fig. 4: The threshold to evaluate the difference between input image and background model is calculated; a binary foreground mask is determined; people are detected at the foreground areas; background model is updated for all non-person areas.

Fig. 5

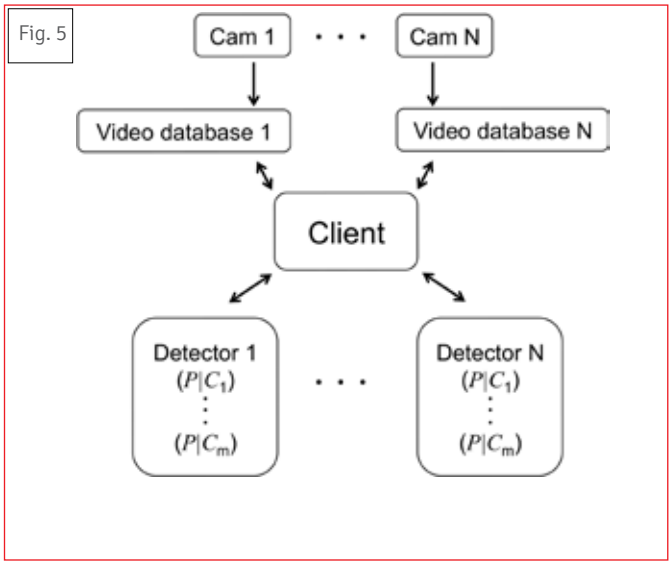

Fig. 5: The camera images are recorded in separate databases. A client can request an image from a database and send it to a detector which in turn detects ROIs within that image and returns these results to the client. Each detector represents a single computer; the tuples (P|C) are parallel HOG iterations.

Fig. 6

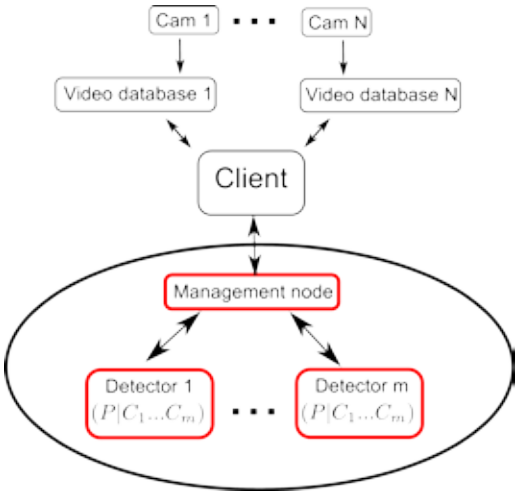

Fig. 6: The image is send to the management node and distributed among the unused detectors. The results will be forwarded to the client by the management node.

of correct detections, the described approach also yielded long computation times (up to 10 seconds for a single HOG iteration). Thus we implemented a GPU based version of the detection algorithm and reduced the computation time to 90 milliseconds, which represents a speedup factor of more than 100. Utilizing multiple GPUs (one for each classifier) in parallel enabled us to finally remove this drawback from our detection framework. In our evaluation of the detector system we dedicated a single computer to each camera, which enabled us to process the camera data in realtime (i. e. 10 fps at 1600px x 1.200px). The systems structure is depicted in Fig.5. Such an approach becomes unfeasible in realistic scenarios, e.g. multiple cameras with high frame rates, as the physical limits of mainstream computers are exceed, since they can't be equipped with enough GPUs. Thus we developed a software framework to distribute the HOG in a cluster-like manner among small computation nodes, each equipped with one or two GPUs (Fig. 6).

Our system structure follows the concept of a Beowulf cluster. The benefits of such structure are the facts that it allows a simple expansion for higher workloads and that small computation nodes can be used. Each node can be equipped with just one GPU, which can be done easily on ordinary computers. The management node will take care of an adequate workload scheduling among the available detector nodes. Furthermore it will provide the client with correctly assembled detection results from the nodes, as a single detection request from the client may be distributed among multiple computers. As the detection is entirely executed on the node's GPU, the CPU resources remain largely available for further preprocessing. In addition to the distributed computation we are planning to reduce the overhead within the parallel execution of HOG iterations. The HOG iterations within a single detector work on the same image data, thus the preprocessing P of the image data can be executed only once. Our described approach does not only allow a flexible system expansion for massive video data streams, it also provides reliable detection results using state of the art people detectors.

Fig. 7: The used features will be extracted from three areas which are located relative to the detection.

Feature Tracking

A feature tracker is used to extract and group first features of the detected people to speed up the later recognition process. To group features of one person's clothes over time, a template based visual tracking of people is used.[4] This fast tracking (> 100 fps) generates long continuous tracks with occlusion handling. Furthermore, subtracks are generated by a feature tracker, which clusters similar features. To generate these subtracks, general features are estimated which are basically described for a human robot dialog system[5]. In this work, the HSV-representation of the color is used which hue and saturation is mostly independent in illumination. The used features are extracted on three fixed areas of the detected people. One rectangle part of the lower body and one of the upper body are separated to determine the mean hue and saturation separately for both areas. At the rectangular upper body part, the mean horizontal and vertical texture rates are calculated with the help of the Scharr filter[6]. The mean horizontal and the mean vertical texture rates describe the texture at the selected area in a compact manner. One 16 bin histogram of the hue and one of the saturation are calculated at an oval area of the upper body. The mean hue and saturation describes the basic color of the user's lower and upper body, while the histograms describe the upper body in detail. By using the hue and saturation histograms, even prints, patches etc. are represented in a very compact form. In this work, normalized histograms are used due to their scale independency.

Merging Image-based Hypotheses[7]

Since multiple detectors (face, upper body, full body) for each camera are used, these different hypotheses are fused into a single person hypothesis. Additionally, hypotheses into overlapping camera views are merged. For this, each hypothesis is transformed into global coordinates which is possible by calibrating each camera.

Centralized components

The APFel project aims to support human operators in analyzing the video data. The operator is able to select a suspicious person on a central control and visualization unit which induces the system to start the search. The results of the camera-based data analysis are stored in the central database. These results are analyzed using a geometrical model of the airport (spatio temporal reasoning[8]) which reduces the huge amount of data enormously. The previously

4. Kolarow, A., Brauckmann, M., Eisenbach, M., Schenk, K., Einhorn, E., Debes, K., Gross, H.-M., Vision-based Hyper-Real-Time Object Tracker for Human-Robot Interaction, in IROS, 2012

5. Hommel, S., Rabie, A., Handmann, U., Attention and Emotion Based Adaption of Dialog Systems, in Intelligent Systems: Models and Applications, vol. 3, Springer Berlin Heidelberg, pp. 215-235, 2013

6. Scharr, H., Optimal operators in digital image processing, Ph.D. thesis, Interdisciplinary Center for Scientific Computer, Ruprecht-Karls-Universität, Heidelberg, 2000

7. Kolarow, A., Schenk, K., Eisenbach, M., Dose, M., Brauckmann, M., Debes, K., Gross, H.-M., APFel: The Intelligent Video Analysis and Surveillance System for Assisting Human Operators, in AVSS, 2013

8. Ibid.

9. Ibid.

10. Eisenbach, M., Kolarow, A., Schenk, K., Debes, K., Gross, H.-M., View Invariant Appearance-based Person Re-identification Using Fast Online Feature Selection and Score Level Fusion, in AVSS, pp. 184-190, 2012

extracted camera-based person features are compared to each other and a first list of hypotheses is created. These restrictions speedup the system enormously, since these enables an analysis of the data in only those time slots which are temporally and geometrically reasonable. This approach allows an effective analysis of the video data and supports the human operator. The time-consuming analysis of the entire data can be avoided. The operator confirms hypotheses which make the overall system more robust.

1 | Prediction[9]
Since it is necessary to search through all the recordings to find sequences containing the person of interest, it saves time to prioritize the processing sequence of the person hypotheses based on statistics for camera transition and abidance times.

2 | Face Recognition
The face recognition component is essential for recognizing a person in a camera network, since face recognition imposes some quality requirements on an image, especially with respect to facial resolution. For this, an additional Appearance-based Re-identification is used to operate without these requirements.

3 | Appearance-based Re-identification[10]
Since the appearance of people's clothes varies significantly, it is important to use a large feature set for re-identification and select discriminative features for a specific person on the fly in the enrollment phase. Using a small subset of well suited features as a template ensures fast matching (12,000 per second).

Conclusion and future work

Through the combination of decentralized and centralized components, the current system is able to track persons in multi camera networks in realtime.

THE SYSTEM HELPS THE SECURITY STUFF AT AIRPORTS TO EVALUATE CRITICAL SITUATIONS VERY FAST AND EXACTLY.

Further works of this project are testing and evaluating the system's usability from the security operator standpoint, and its acceptance by persons who use the airport infrastructure, such as pilots, passengers, and airport employees. Furthermore, in order to satisfy the need for speed in security applications, the algorithms will be parallelized. Current work of this project is the tracking of single persons within larger groups of people, as well as tracking whole groups of people. Although the system is primarily intended to be used at airports, it is possible to adapt the system to other public areas, like train stations or subways.

Acknowledgment:

This work was funded by the German Federal Ministry of Education and Research (BMBF).

In 2004, SEBASTIAN HOMMEL started his study of computer science at the TU-Ilmenau. During his studies Hommel deepened the medical computer science, and the neuroinformatics and cognitive robotics. He completed his diploma thesis „Zeitliche Analyse von Emotionen auf Basis von Active Appearance Modellen“ at Neuroinformatics and Cognitive Robotics Lab of TU-Ilmenau in 2010.

Afterwards, Hommel joined the Computer Science Institute of the University of Applied Sciences Hochschule Ruhr West as a research assistant. He started his PhD study in cooperation with the University of Duisburg-Essen in the field of texture and color based recognition in 2012.

MATTHIAS A. GRIMM studied Applied Computer Science at the Ruhr-Universität Bochum and completed his studies with a thesis about behavioral organization for mobile robotic systems. Afterwards, he started to work as a research assistant at the Computer Science Institute of the University of Applied Sciences Ruhr West with the focus on image processing. Since 2012, he is a PhD student in cooperation with the University of Duisburg-Essen and works in the field of gait recognition for people identification.

DARIUS MALYSIAK studied electrical engineering at FH-Bochum and concluded his study with a thesis about data redundancies within algorithms for face recognition. Afterwards, he continued with the studies of computer science and mathematics at the Ruhr-Universität Bochum, which he finished with a thesis about machine learning approaches for side-channel attacks and generic decoding algorithms for multiple syndromes, respectively.

In 2010, Malysiak joined the Computer Science Institute of the University of Applied Sciences Hochschule Ruhr West as a research assistant. Since 2012, he is a PhD student at the Ruhr-Universität Bochum and researches efficient algorithms in the area of machine learning.

In 1995, PROF. DR.-ING. UWE HANDMANN joined the Institut für Neuroinformatik, Chair of Theoretical Biology as a scientific staff member. The Institut für Neuroinformatik (INI) is an independent research institute at the Ruhr-Universität Bochum, Germany. Between 2000 and 2008 he was employed by L1 Identity Solutions AG. There, he was responsible for video based face recognition systems. In 2008 , Dr. Handmann took up a position as Professor at the University of Applied Sciences Südwestfalen. He was leading the Laboratory for Computer Science and Media Technology. In 2010, Dr. Handmann joined the University of Applied Scieces Ruhr West (Hochschule Ruhr West). He is Head of the Computer Science Institute (Institut Informatik) and is responsible for several research projects in the field of video processing and intelligent systems.

PRESENTATION OF THE INSTITUTE

The Hochschule Ruhr West (HRW) – University of Applied Sciences – is a young university (founded in 2009) which is located in Bottrop and Mülheim an der Ruhr. Research is conducted in a total of seven institutes. At the HRW, the APFel project is executed from the Computer Science Institute. The core assignment of this institute is teaching and research in the subject of Applied Computer Science.

Applied computer science is regarded as an interdisciplinary interface between information science and other scientific disciplines, particularly technology and the natural sciences. The key areas of research are Vehicle Information Technology and Cognitive System Technology.

The overall aim of Cognitive System Technology is the implementation of intelligent systems in private households and industrial automation (e. g. for inspection tasks or quality assurance) that are able to react independently to interaction partners while processing varying ambient conditions in a targeted manner. The second research focus of the institute is Vehicle Information Technology. In addition to exploring the use of sensors and actuators in motor vehicles and vehicular data processing, various projects center on research activities in the area of intelligent assistance systems.

The Computer Science Institute is divided into several labs which cover the whole spectrum of applied computer science by a close collaboration. The APFel project is mainly located at the laboratory for Image Processing and Neuroinformatics (Prof. Dr.-Ing. Uwe Handmann). This lab is focused on the analysis of image contents in real-time, biologically inspired methods, intelligent vehicles and advanced driver assistance systems, intelligent video technology, person identification, sensor data fusion and psychoacoustics.

www.institut-informatik.de
www.computer-science-institute.de
www.hochschule-ruhr-west.de

PROJECT PARTNER

L-1 Identity Solutions AG
(area of responsibility: project coordinator; multi-camera facial people recognition and tracking; spatio-temporal reasoning)
Universitätsstr. 160, 44801 Bochum
www.morpho.com

TU-Ilmenau, FG Neuroinformatik und Kognitive Robotik
(area of responsibility: multi-camera full-body people detection and tracking; movement prediction and statistics; appearance-based re-identification)
Helmholtz-Platz 5, 98693 Ilmenau
www.tu-ilmenau.de/neurob

Hochschule Ruhr West
(area of responsibility: system specification and evaluation; video indexing; sensor control)
Prof. Dr.-Ing. Uwe Handmann
Tannenstr. 43, 46240 Bottrop
uwe.handmann@hs-ruhrwest.de

Ruhr-Universität Bochum
(area of responsibility: legal accompanying research)
Postfach 10 21 48, 44801 Bochum
thomas.feltes@rub.de

Avistra GmbH
(area of responsibility: acceptance analysis; probability of presence at several times)
Zum Langen See 39, 12557 Berlin
radig@avistra.de

Flughafen Erfurt-Weimar GmbH
(area of responsibility: end user; test environments)
Binderslebener Landstr. 100, 99092 Erfurt
matthias.koehn@flughafen-erfurt.de

European Aviation Security Center e.V.
(area of responsibility: end user; test environments; acceptance analysis)
Am Flugplatz, 14959 Schönhagen
info@eascschoenhagen.org