Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall

Henri Bouma *, Jan Baan, Sander Landsmeer, Chris Kruszynski,
Gert van Antwerpen, Judith Dijk

TNO, P.O. Box 96864, 2509 JG The Hague, The Netherlands

ABSTRACT

The capability to track individuals in CCTV cameras is important for e.g. surveillance applications at large areas such as train stations, airports and shopping centers. However, it is laborious to track and trace people over multiple cameras. In this paper, we present a system for real-time tracking and fast interactive retrieval of persons in video streams from multiple static surveillance cameras. This system is demonstrated in a shopping mall, where the cameras are positioned without overlapping fields-of-view and have different lighting conditions. The results show that the system allows an operator to find the origin or destination of a person more efficiently. The misses are reduced with 37%, which is a significant improvement.

Keywords: Surveillance, tracking, CCTV, person re-identification, multi-sensor fusion.

1. INTRODUCTION

The capability to track individuals in CCTV cameras is important for surveillance applications at e.g. train stations, airports and shopping centers. For the camera operators, however, it is laborious to track and trace people over multiple cameras. In this paper, we present a semi-autonomous system for real-time tracking and fast interactive retrieval of persons in video streams from multiple surveillance cameras. This system is demonstrated in a shopping mall. We describe our system, which consists of tracklet generation, re-identification and a graphical man-machine interface. The system is tested in a shopping mall with eight static cameras and 6 cameras were selected for an operator-efficiency experiment. These cameras have non-overlapping (or hardly overlapping) field-of-views and different lighting conditions. All video streams are processed in parallel on a distributed system and tracks and detections are continuously stored in a database. The operator can use these tracks and detections to quickly answer questions such as "where did a particular person come from?" or "where did he go to?". The interface enables live tracking in current streams and interactive searches in historic data. The results show that the system allows an operator to find the origin or destination of a person more efficiently with less misses.

The outline of this paper is as follows. Section 2 describes our system, Section 3 describes the experiments and results and finally Section 4 summarizes the conclusions.

2. METHOD

The system overview is shown in Figure 1. The main components are tracklet generation and the re-identification engine and a graphical man-machine interface. Tracklet generation is an activity that continuously processes the incoming video streams to detect persons and track them within a single camera. The resulting tracklets are stored in a tracklet database. This database allows our system to quickly retrieve similar candidates after human interaction without computational intensive video processing. In order to track a person in a large environment over multiple non-overlapping cameras, the separate tracklets of a certain person from different cameras need to be combined. The re-identification engine compares

* henri.bouma@tno.nl; phone +31 888 66 4054; http://www.tno.nl

Henri Bouma, Jan Baan, Sander Landsmeer, Chris Kruszynski, Gert Van Antwerpen, Judith Dijk, "Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall", Proc. SPIE, Vol. 8756, (2013).

the query with tracklets in the database and presents the most likely candidates. This engine consists of two components: appearance-based matching and space-time localization. The combination of both is used to present the best matching candidate. The human-machine interface allows the operator to interact with the system, by selecting queries and candidates. Each component is described in more detail in the following subsections.

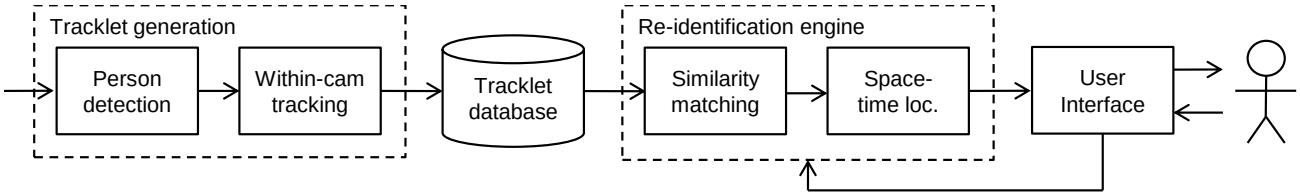

Figure 1: The system consists of tracklet generation, a re-identification engine and a graphical user interface.

Our video-processing framework consists of robust re-usable video-processing components that can handle different types of inputs (e.g. standard IP cameras), flexible parallel and distributed processing over multiple computers in a network, and reliable data transfer by using FIFO buffers in the real-time processing environment.

2.1 Person detection

Background subtraction is commonly used for the detection of moving objects, but it typically fails to reliably segment complete persons in busy environments [25]. Therefore, we used a pedestrian detector that can recognize humans in a static frame. There are many pedestrian detectors available in literature, and we used several to detect persons in the video, such as: the Laptev detector [22], Felzenszwalb detector [15], GPU based Felzenszwalb, Cascaded Felzenszwalb [16] and the FPDW [11].

2.2 Tracklet generation

In literature, others have tried to solve the problem of multi-target tracking. Poiesi e.a. [24] proposed a method for multi-target track-before-detect based on particle filtering. They included the target ID into the particle state to handle unknown and large number of targets. A Markov random field (MRF) is used for death and birth of targets, assignment of IDs to targets is performed using mean-shift clustering supported by a Gaussian mixture model. Heili e.a. [19] used a detection-based approach for multi-person tracking with a conditional random field (CRF) for statistical labeling. They used the time gap, Euclidean position difference, and Bhattacharyya color distance to model similarities and dissimilarities, and learn the parameters in an unsupervised way. Heili e.a. first apply their tracking algorithm using small temporal association windows. Then, the features are recomputed and the parameters are relearned up to the desired association windows. They also propose to make the parameters dependent on the local context, e.g. if the local space-time region of a video is crowded, then the distances between detections in this region will be smaller, and the model shrinks the spread of feature distributions. Englebienne e.a. [13][14] used a probabilistic approach to track people across multiple, sparsely distributed cameras, where an observation corresponds to a person walking through the field of view of a camera. Modeling appearance and spatio-temporal aspects probabilistically allows them to deal with uncertainty. They performed an experiment with two stereo cameras and four volunteers where the fields of view overlap slightly. The probability of the appearance is assumed to be normal and additive, the probability of transiting from one camera to another is set beforehand, and the probability of travel time is modeled as a uniform distribution based on two conditions. Barbu e.a. [1] proposes an approach that involves simultaneous object detection and tracking, which optimizes the joint object-detection and temporal-coherency score. Many detectors, e.g. those of Felzenszwalb, use internally a scale-space pyramid to represent all possible detections at all locations and scales in an associated frame. Instead of extracting and tracking the thresholded detections, Barbu e.a. directly track all detections in the entire pyramid simultaneously by defining a distance measure between detection pyramids for adjacent frames and performing the Viterbi tracking algorithm on these pyramids. Employing a distance transform makes this process linear in the number of location and scale positions in the pyramid. The simultaneous detection and tracking solves the missing detections in

single frames, which could also be solved by lowering detection thresholds or by projecting detections forward to augment the raw detector in subsequent frames. Hu e.a. [21] show an extensive overview of tracking algorithms for crowded scenes and they proposed a novel tracking method for crowded scenes based on particle filtering and optic flow.

The tracking that is used in this paper is based on an detection-before-tracking approach and it consists of a prediction and an association step. The prediction (and update) step has been implemented in our framework with a linear fit through the last *N* points (which assumes constant velocity). The association step is based on the bounding-box overlap between the predicted and the actual location of a detection. New tracks, which have a less reliable velocity estimate, can also connect detections in the vicinity without overlap. New tracks are not created in a location without moving objects, to avoid the inclusion of false static detections, such as a garbage bin. However, the component is implemented in such a way that it will track pedestrians that loiter or stop after entering the scene. The fitness of a track is related to the time when the last detection was added; when no novel detections have been added for a few seconds, the track dies.

2.3 Distribution and storage of tracklets

Our hardware setup consists of one master PC for central database purposes, multiple slave PCs for distributed tracklet generation in multiple cameras, and one machine for the graphical user interface (GUI). The tracklet generation is executed and logged at multiple machines, and the tracklets are distributed to the master machine and external clients with JSON and httppost. Recent tracklets are centrally cached in the memory on our master machine and also on our GUI machine. The current tracklets can be retrieved as a continuous stream of information and historic tracklets can be retrieved by providing a start and end time. Each tracklet consists of the following fields: a unique track number, camera number, bounding box (x,y,width,height in image coordinates), positions (x,y,z in meters), date and time (ISO UTC: yyyymmddThhmmss.uuuuuu), a snippet (small image) for the GUI, and a signature for re-identification. Furthermore, we have couple messages based on the re-identification described in section 2.4 to connect tracklets from different cameras, which contain a unique couple number and the related track numbers.

2.4 Appearance based re-identification

Recently, Satta e.a. [26] proposed a dissimilarity-based approach for speeding up existing re-identification methods and an approach for retrieving images of individuals based on a textual query describing clothing appearance, instead of an image. The same authors [27] proposed an online re-identification system in a camera network which uses segmentation based on the depth map of a Kinect.

In our system, selection of similar tracklets uses a signature for each tracklet. These signatures are based on the re-identification histograms [4], which are already computed during tracklet generation. The selection of the detection on which the signature is computed is based on the largest product of bounding-box height and detection confidence. Improved matching can be obtained when multiple signatures are used from a track, e.g. by using all histograms or by selecting one per cluster of similar histograms. The signatures are used to compute the similarity between a tracklet in one camera and all other tracklets.

The description of the multi color-height histograms (MCHH) and the transformed-normalized RGB histograms – which are used for appearance-based similarity – can be found in [4]. Several detailed implementation notes about the computation of these histograms – which are important for reproducibility and missing in the previous paper – are the following. The first note is related to the segmentation of the person that is more accurate than the rectangular bounding-box of the person to avoid the influence of background pixels. Initially, we used a fixed segmentation mask in the pedestrian detection, but recently it was replaced by a segmentation based on background subtraction. Only values inside the binary person mask are used for the normalization (RGB-rank, transformed, normalized). The other notes are about computation of the transformed-normalized RGB. The transformed color-space ranges from -2 sigma to +2 sigma and the region in the tails of the distribution is collected in the first and last bin. The normalized color-space has a normalized *r* and *g* component and the *b* collects the total intensity (*b* = R+B+G). Transformed-normalized RGB is implemented in the following ordering: normalized(transformed(RGB)).

2.5 Space-time localization

A person cannot be at two different locations at the same time and the maximal travelling speed is used to speed up the process of finding similar people by eliminating tracklets that are too far away.

The distance between two points in the shopping mall is not computed as an Euclidean straight-line distance through all buildings, but as the geodesic distance through areas that are publically accessible. The start- and endpoint of a track are projected to a graph that is centered at corridors and the travel time is computed with the graph assuming an average traveling speed.

2.6 Visualization and user-interaction

The graphical user interface (GUI) consists of several panels, as shown in Figure 2: a *map* panel, a *spot-camera* panel, an *other-cameras* panel, and a *candidate-selection* panel. The map panel consists of a map on which the camera locations and the generated tracks are projected. The spot camera, i.e. one camera with high resolution, can be selected either by interacting with the other cameras, with the map, or with the candidate-selection region. The interface can show both live video streams or pre-recorded videos. The user can interact with this spot-camera view by playing forward or backward and by selecting a person of interest. The panel with a flexible number of other cameras allows to get a quick overview and to select the spot camera. The candidate-selection panel has two axes: the horizontal axis shows the time and the vertical axis shows the different cameras. All candidates (generated by tracklet generation) can be shown in this panel. When a person of interest is selected (either in the spot-camera view or in the candidate view), false candidates can be removed based on space-time localization or on appearance-based similarity. The user is able to interact with the similarity threshold to show more or less candidates (see Figure 3). By selecting the correct candidates, a complete track of an individual can quickly be generated. This track is also visualized in the map panel, the spot-camera panel and the candidate-selection panel (Figure 3).

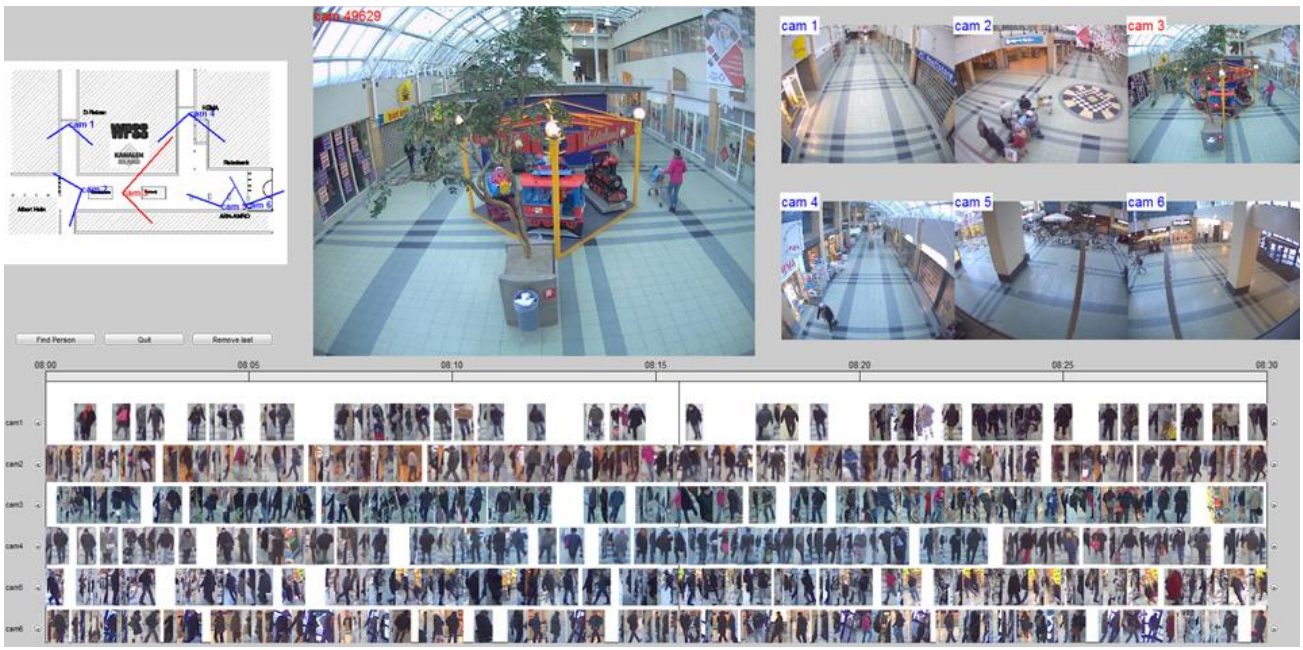

Figure 2: The graphical user interface includes the following: map (top-left), spot-camera view (top-center), other cameras (top-right), candidate view (bottom). The spot-view camera is displayed in red on the map. The candidate view displays time on the horizontal axis and cameras on the vertical axis.

Figure 3: When searching using the re-identification engine, many candidates are suppressed and only people with similar appearance are shown. The line in the candidate panel connects the tracklets of a certain person.

3. EXPERIMENT AND RESULTS

3.1 Experimental setup

In the shopping mall, nine PCs and eight static cameras were installed to test our real-time tracking and re-identification framework. In the operator-efficiency experiment (Sec. 3.4), only six of the cameras were used to create more blind spots. The PCs are Dell Optiplex 9010 small form, with Intel Core i7-3770, 3.4 GHz, and 8GB DDR3 memory. There are different types of cameras, including AXIS 211M network cameras with 1280x1024 resolution.

3.2 Detection and tracklet generation results

To select a detector in our tracklet generation, we compared several implementations of Felzenszwalb (CPU, GPU, Cascaded) and the FPDW. The FPDW appeared to give good results on the high-resolution images in acceptable computational time. An example of the detection results with the FPDW detector is shown in Figure 4.

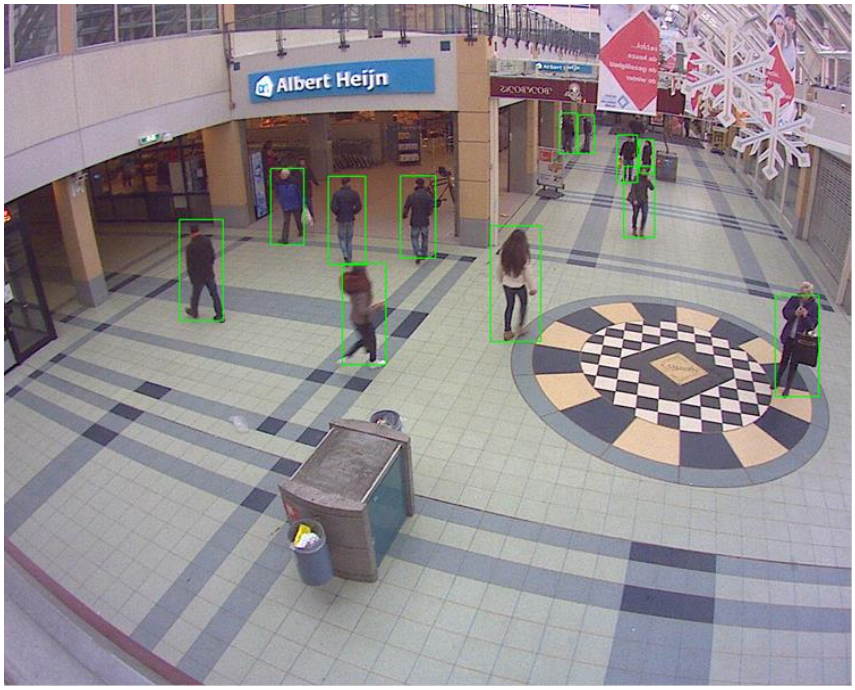

Figure 4: Example of detection results with the FPDW detector. The green boxes indicate the detections.

Tracklet generation consists of pedestrian detection and within-camera tracking. An example of generated tracklets in a camera view is shown in Figure 5 and a visualization on the map is shown in Figure 6.

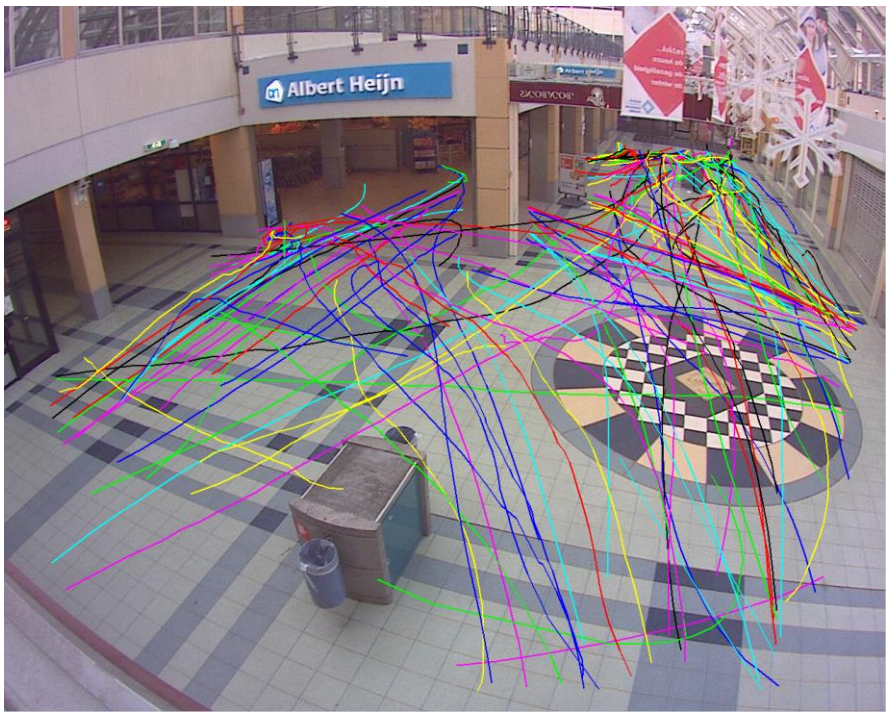

Figure 5: Example of generated tracklets in approximately 2 minutes. All lines indicate different tracklets, i.e. different persons.

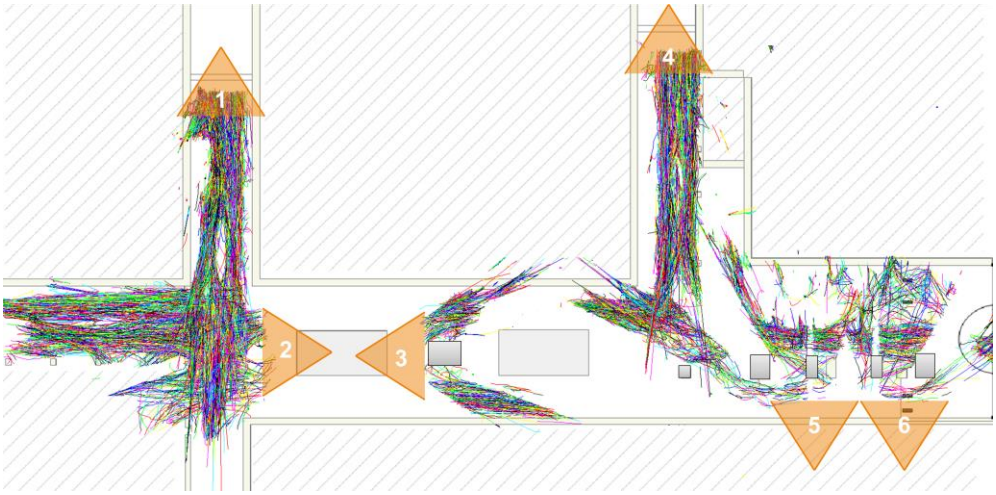

Figure 6: Map of a region in the shopping mall with several cameras (orange triangles) and tracklets (colorful lines) in approximately 6 minutes.

3.3 Matching results

The experiment to assess the quality of different color models was performed on person pairs from different cameras in the shopping mall (77 person pairs) and on data from the VIPeR dataset (632 person pairs) [18]. The color models and parameters were selected from [4] and several combinations were tested.

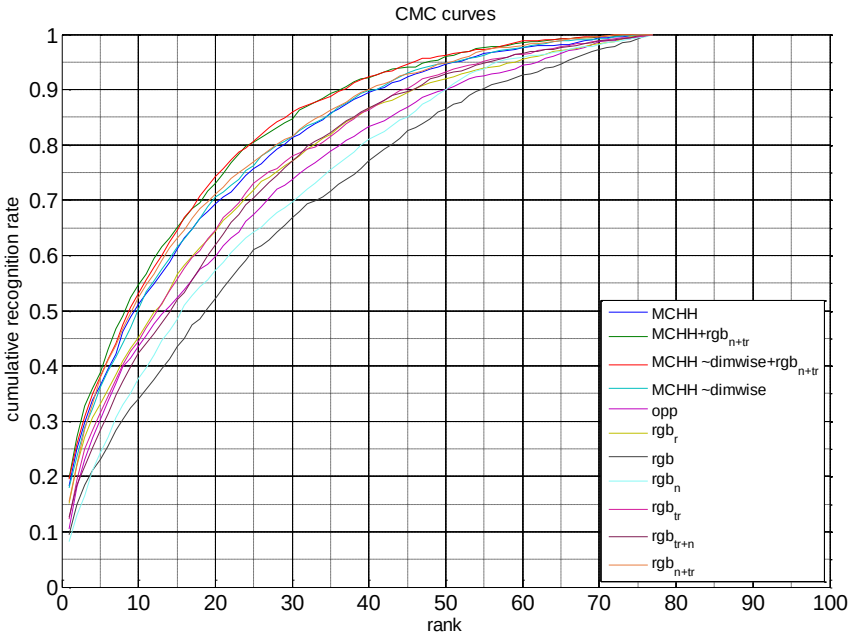

Figure 7: CMC of different color models on data from the shopping center (77 person pairs): multi color-height histograms (MCHH), transformed (tr), normalized (n), opponent (opp), and RGB-rank (r). The figure shows that a combination of MCHH with transformed-normalized RGB results in the best performance.

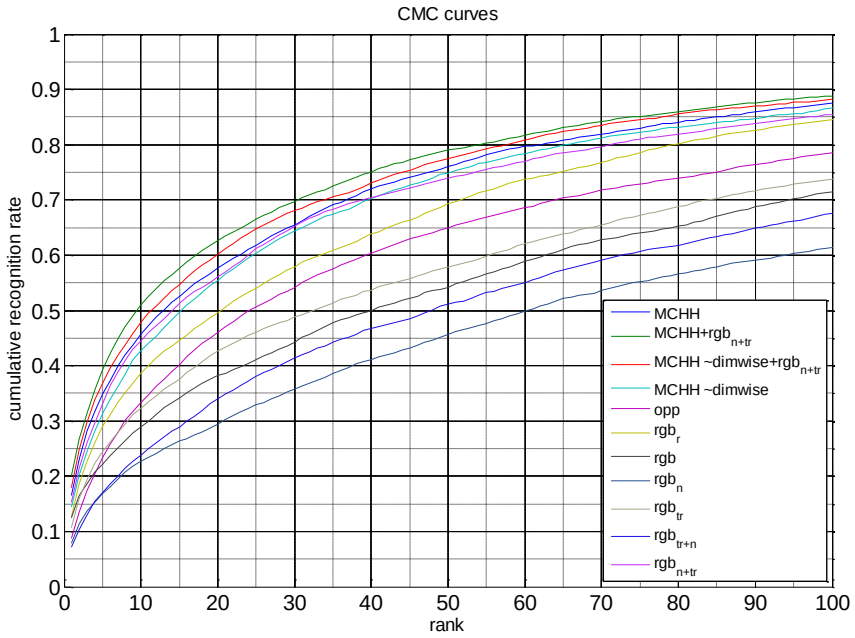

Figure 8: CMC of different color models on data from VIPeR [18] (10 randomly selected sample sets of 316 person pairs out of the 632 pairs): multi color-height histograms (MCHH), transformed (tr), normalized (n), RGB (rgb), opponent (opp), and RGB-rank (rgb_r). The figure shows that a combination of MCHH with transformed-normalized RGB results in the best performance.

The results are shown in Figure 7 and Figure 8. The figures show similar results on VIPeR data and in the shopping mall. Furthermore, the combination of MCHH with transformed-normalized RGB (both described in [4]) performs even better than each of them separately.

The size of the database has a large effect on the matching performance. In a larger database, it is likely to have a larger number of images with a high similarity to the query image, which do not actually constitute a correct match. In this experiment, performed on the shopping mall data, the standard set of database images was enlarged by adding a varying numbers of images from another period than the standard images.

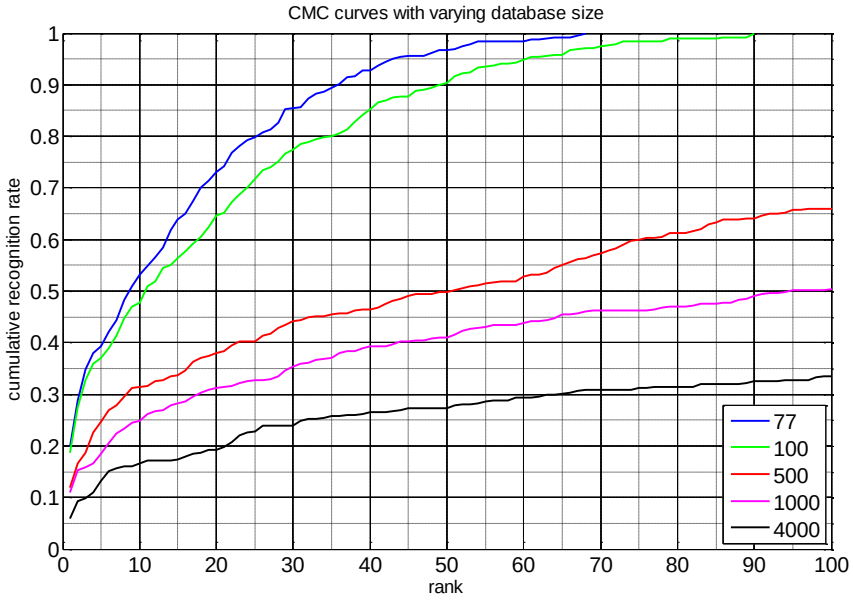

Figure 9: CMC curves for varying database sizes, where the rank on the horizontal axis is an absolute value.

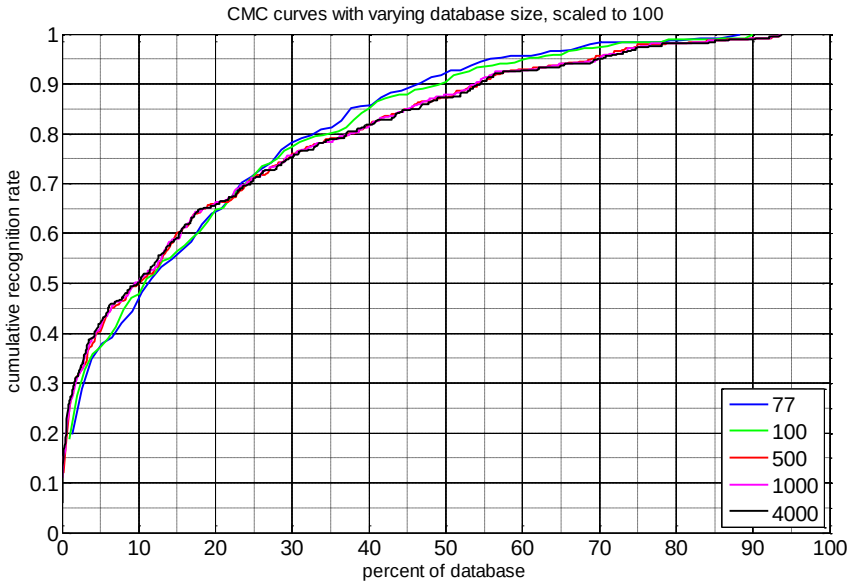

Figure 10: CMC curves for varying database sizes, where the percentage on the horizontal axis is the rank relative to the database size.

Figure 9 shows that the CMC curve decreases for larger database sizes, as expected. The results in Figure 10 show that the CMC remains approximately constant on different database sizes, when the horizontal axis is rescaled to 100% of the database size (instead of the absolute value of the database size on the horizontal axis). This allows us to estimate how well the system will perform in sparse or crowded environments.

3.4 Operator efficiency results

Finally, we performed an operator-efficiency experiment to test how good a human operator can track persons over multiple cameras with and without our proposed system. The core of the proposed system is the candidate view which allows users to interact with the tracklets and use the re-identification engine. In both cases – with and without candidate panel – the volunteers were allowed to use the view with synchronized cameras, the time-bar, the map and the panel with the spot camera, although some of these elements may already be an improvement over commonly used surveillance systems.

In the experiment, we selected pedestrians (N=17), six cameras and for several pedestrians a time slot of 30 of pre-recorded video data and for other pedestrians a time slot of 45 minutes. Many of the pedestrians are entering the scene more than once, e.g. after some time in a shop. The pedestrians had different appearances, several with bright colors (e.g., pink, red, blue) and others with more common colors (e.g. gray, black, brown). For each pedestrian the following was given as a query: an initial camera number, a time stamp that the pedestrian is visible in this camera, a complete camera frame with a red bounding box indicating the location of the pedestrian, and an enlarged snippet of the pedestrian to show the appearance. The tool was initiated at the correct initial query position (time and camera) for each pedestrian.

The search task was performed by volunteers (N=8) that were trained for approximately 15 minutes to interact with our GUI. After the training, the volunteers were given the instruction to find the appearance of the query persons in other cameras as good as possible within max. 5 minutes. Each volunteer had to track half of the pedestrians without the candidate view and the other half with the candidate view. Half of the volunteers was presented the odd pedestrians with the candidate view and the even without, and for the other half of the volunteers it was vice versa. The alternating scheme was chosen to avoid effects of a learning curve. The order of the presented pedestrians was fixed during the experiment. One volunteer was removed from the experiment and replaced by a ninth volunteer, because the instruction appeared to be misunderstood (pedestrians were only tracked forward and not backward, which made this volunteer almost twice as fast as the other operators but at the cost of more FN).

The hits (TP), false positives (FP) and misses (FN) have been checked by visual inspection. The central question that the volunteers had to answer is: 'Where did the pedestrian come from and where did he/she go to?'. Therefore, the start and the end of a track are the most important. One mouse-click indicating the presence of a pedestrian is sufficient for a camera, unless a pedestrian reappears later after leaving the camera view, e.g. by leaving the shopping mall or entering a shop. For example, a pedestrian may enter the mall in camera 1, move quickly via camera 2 to camera 3, enter a shop in camera 3, and reappear 5 minutes later in camera 3 and leave the mall in camera 4. In this case, the volunteer should have indicated the presence in camera 1 (once), camera 3 (twice: before and after the shop) and camera 4 (once). A response in camera 2 is ignored, because it is a trivial intermediate point. The scoring system was explained to the volunteers before they started to track the pedestrians.

In total 78 possible detections (TP + FN) were checked (on average 4.6 per pedestrian), of which TP+FN=30 in the even pedestrians and 48 in the odd pedestrians. The number of false positives was very small (on average 0.6 FP per operator for all 17 pedestrians). There is no significant difference in timing, because most volunteers used the available five minutes completely (on average 265 and 250 seconds without and with candidates respectively). Therefore, we evaluate the FNs. The results are shown in Table 1 (for each volunteer) and Table 2 (for each pedestrian). Because of the difference in total possible misses between the even and odd pedestrians, also a relative number of FN/(TP+FN) was computed in Table 1. Both tables show that the number of misses (FN) is reduced with 37% when the system is used. This reduction in FN is significant using the Sign Test (in Table 2 the number of non-ties = 16, $p = 0.003$).

Table 1: Comparison of operator efficiency with and without our system based on 8 volunteers who received 17 query pedestrians. The absolute and relative number of FN are shown for each volunteer. The results show a reduction of 37% of the misses when the system is used.

	Pedestrians with candidate view	**Absolute number of FN: FN**		**Relative number of FN: FN / (TP + FN)**	
		w.o. cand.	**with cand.**	**w.o. cand.**	**with cand.**
Operator 1	even	29	10	0.60	0.33
Operator 2	odd	13	18	0.43	0.38
Operator 3	even	19	6	0.40	0.20
Operator 4	odd	12	17	0.40	0.35
Operator 5	even	25	9	0.52	0.30
Operator 6	odd	13	13	0.43	0.27
Operator 7	even	31	8	0.65	0.27
Operator 8	odd	9	14	0.30	0.29
Average FN		18.9	11.9	0.47	0.30

Table 2: Comparison of operator efficiency with and without our system based on 8 volunteers who received 17 query pedestrians. The absolute number of FN are shown for each pedestrian. The results show a reduction of 37% of the misses when the system is used.

Pedestrian number	**Odd**									**Even**								**Total**
	1	**3**	**5**	**7**	**9**	**11**	**13**	**15**	**17**	**2**	**4**	**6**	**8**	**10**	**12**	**14**	**16**	
Possible detections: FN+TP	4	4	6	2	5	8	8	5	6	3	2	2	3	4	4	6	6	78
FN without candidate view	9	7	15	2	12	22	16	9	12	3	1	0	4	7	3	16	13	151
FN with candidate view	3	7	11	0	1	10	13	8	9	5	1	0	3	2	2	12	8	95
Difference in FN	**6**	**0**	**4**	**2**	**11**	**12**	**3**	**1**	**3**	**-2**	**0**	**0**	**1**	**5**	**1**	**4**	**5**	**56**

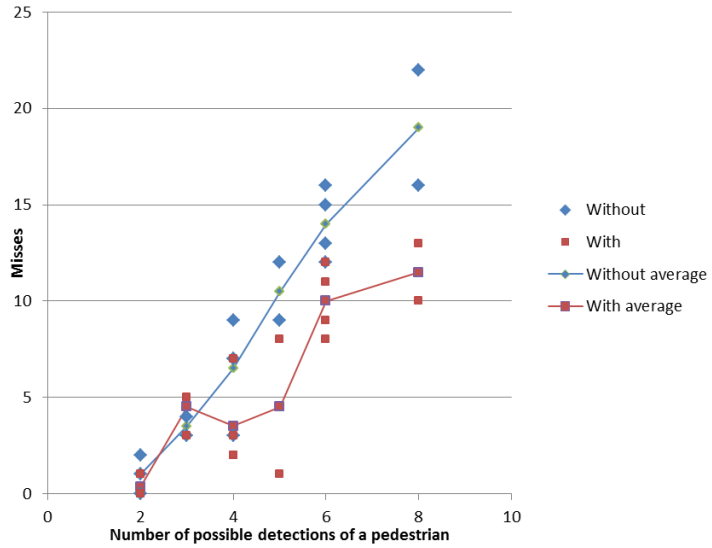

Figure 11: Missed pedestrians with and without candidate view, as a function of the number of possible detections. It is shown that the results for the system especially improve for a higher number of possible detections of a pedestrian.

The results indicate that the improvement is primarily found for the pedestrians which appear more than 3 or 4 times in the scene. This effect is also shown in Figure 11. Therefore, we tested the significance of the improvement for a small number of possible detections (smaller or equal than 4) and for pedestrians with a high number of possible detections (larger than 4). With the Sign Test, it was found that the improvement using the system was not significant for the first set (number of non-ties 6, $p = 0.22$). The improvement for the pedestrians with a high number of possible detections was found to be significant (number of non-ties = 8, $p = 0.013$). This shows that the improvement of the system is mainly caused by persons who appear more often in the scene.

4. CONCLUSIONS AND DISCUSSION

In this paper, we presented a semi-autonomous system for real-time tracking and fast interactive retrieval of persons in video streams from multiple surveillance cameras in a shopping mall. We described our system, which consists of pedestrian detection, track generation, space-time localization, appearance based re-identification and a graphical man-machine interface to fuse the information from multiple cameras. The system was tested in a shopping mall with multiple static cameras. These cameras have non-overlapping field-of-views and different lighting conditions. All video streams are processed in parallel on a distributed system and tracks and detections are continuously stored in a database. The operator can use these tracks and detections to quickly answer questions such as "where did a particular person come from?" or "where did he go to?". The interface enables live tracking in current streams and interactive searches in historic data. The results show that the system allows an operator to find the origin or destination of a person more efficiently with 37% less misses, which is a significant reduction. It was also found that the reduction was mainly found for persons who appeared more often in the scene.

The current experiment was relatively small in time (max. video duration = 45 minutes), number of cameras (6) and space (scene contained only a few blind spots with these 6 cameras). Yet, the results already showed an improvement of 37%. The volunteers indicated after the operator-efficiency experiment that they experienced most added value of the candidate view when a pedestrian was outside the visible region for a long time. We expect that a large scale experiment (larger scene, more cameras, more blind spots and pedestrians out-of-view for long duration) will even show a higher performance gain.

Future work may include a large scale experiment, the combination with behavioral profiling [6] or automatic action recognition [5][7][9], and an evaluation on a public available database for multi-camera surveillance for person re-identification, such as the database of Bialkowski e.a. [2]. Their database consists of 150 sequences of subjects travelling in a building environment with eight camera views, with various viewing angles and illumination conditions, and an XML-based evaluation protocol.

ACKNOWLEDGEMENT

The work for this paper was supported by the 'Maatschappelijke Innovatie Agenda - Veiligheid' in the project: "Watching People Security Services" (WPSS). This project is a collaboration between TNO, Eagle Vision, Vicar Vision, Noldus IT, Cameramanager.com and Borking Consultancy. This consortium acknowledges the "Centrum voor Innovatie en Veiligheid" (CIV) and the "Diensten Centrum Beveiliging" (DCB) in Utrecht for providing the fieldlab facilities and support. The development of the demonstrator was partially funded by the EU FP7-Security project PROTECTRAIL.

REFERENCES

[1] Barbu, A., Michaux, A., Narayanaswamy, S., Siskind, J.M., "Simultaneous object detection, tracking and event recognition," Advances in Cognitive Systems, (2012).
[2] Bialkowski, A., Denman, S., Sridharan, S., Fookes, C., "A database for person re-identification in multi-camera surveillance networks," IEEE Int. Conf. Digital Image Computing Techniques and Appl. DICTA, (2012).
[3] Benenson, R., Mathias, M., Timofte, R., Van Gool, L., "Pedestrian detection at 100 frames per second," IEEE Conf. Computer Vision and Pattern Recognition CVPR, (2012).

[4] Bouma, H., Borsboom, S., Hollander, R. den, Landsmeer, S., Worring, M., "Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination," Proc. SPIE 8359, (2012).
[5] Bouma, H., Hanckmann, P., Marck, J.W., Penning, L., Hollander, R., Hove, J.M. ten, Broek, S.P. van den, Schutte, K., Burghouts, G., "Automatic human action recognition in a scene from visual inputs," Proc. SPIE 8388, (2012).
[6] Bouma, H., Vogels, J., Aarts, O., Kruszynski, C., Wijn, R., Burghouts, G., "Behavioral profiling in CCTV cameras by combining multiple subtle suspicious observations of different surveillance operators," Proc. SPIE 8745, (2013).
[7] Bouma, H., Burghouts, G., Penning, L. de, Hanckmann, P., Hove, J.M., Korzec, S., Kruithof, M., Landsmeer, S., Leeuwen, C. van, Broek, S. van den, Halma, A., Hollander, R. den, Schutte, K., "Recognition and localization of relevant human behavior in videos," Proc. SPIE 8711, (2013).
[8] Bredereck, M., Jiang, X., Korner, M., Denzler, J., "Data association for multi-object tracking-by-detection in multi-camera networks," Proc. IEEE Int. Conf. Distributed Smart Cameras ICDSC, (2012).
[9] Burghouts, G.J., Penning, L., Hove, J.M., Landsmeer, S., Broek, S.P., Hollander, R. den, Hanckmann, P., Kruithof, M., Leeuwen, C., Korzec, S., Bouma, H., Schutte, K., "A search engine for retrieval and inspection of events with 48 human actions in realistic videos," Int. Conf. Pattern Recognition Applications and Methods ICPRAM, (2013).
[10] Dijk, J., Rieter-Barrell, Y., Rest, J. van, Bouma, H., "Intelligent sensor networks for surveillance," Journal of Police Studies: Technology-Led Policing 3(20), 109-125 (2011).
[11] Dollar, P., Belongie, S., Perona, P., "The fastest pedestrian detector in the west," British Machine Vision Conf. BMVC, (2010).
[12] Dollar, P., Wojek, C., Schiele, B., Perona, P., "Pedestrian detection: An evaluation of the state of the art," IEEE Trans. Pattern Analysis and Machine Intelligence 34(4), 743-761 (2012).
[13] Englebienne, G., Oosterhout, T., Krose, B., "Tracking in sparse multi-camera setups using stereo vision," Proc. ACM/IEEE Int. Conf. Distributed Smart Cameras, (2009).
[14] Englebienne, G., Krose, B., "Fast Bayesian people detection," Benelux AI conference BNAIC, (2010).
[15] Felzenszwalb, P., Girshick, R., McAllester, D., Ramanan, D., "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Analysis and Machine Intelligence 32(9), 1627-1645 (2010).
[16] Felzenszwalb, P., Girshick, R., McAllester, D., "Cascade object detection with deformable part models," IEEE Conf. Computer Vision and Pattern Recognition, 2241 – 2248 (2010).
[17] Feris, R., Datta, A., Pankanti, S., Sun, M., "Boosting object detection performance in crowded surveillance videos," Appl. Computer Vision, 427-432 (2013).
[18] Gray, D., Brennan, S., Tao, H., "Evaluating appearance models for recognition, reacquisition, and tracking," IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance PETS, (2007).
[19] Heili, A., Odobez, J-M., "Parameter estimation and contextual adaptation for a multi-object tracking CRF model," IEEE Workshop Performance Evaluation of Tracking and Surveillance PETS, (2013).
[20] Hassan, W., Bangalore, N., Birch, P., Young, R., Chatwin, C., "An adaptive sample count particle filter," Computer Vision and Image Understanding CVIU 116(12), 1208-1222 (2012).
[21] Hu, N., Bouma, H., Worring, M., "Tracking individuals in surveillance video of a high-density crowd," Proc. SPIE 8399, (2012).
[22] Laptev, I., "Improving object detection with boosted histograms," Image and Vision Computing 27(5), 535-544 (2009).
[23] Liu, Y., Shao, Y., Sun, F., "Person re-identification based on visual saliency," Int. Conf. Intell. Systems Design and Applications ISDA, 884 – 889 (2012).
[24] Poiesi, F., Mazzon, R., Cavallaro, A., "Multi-target tracking on confidence maps: An application to people tracking," Computer Vision and Image Understanding CVIU, (2013).
[25] Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B., "Image change detection algorithms: a systematic survey," IEEE Trans. Image Processing 14(3), 294-307 (2005).
[26] Satta, R., Fumera, G., Roli, F. , "Fast person re-identification based on dissimilarity representations," Pattern Recognition Letters 33(14), 1838-1848 (2012).
[27] Satta, R., Pala, F., Fumera, G., Roli, F., "Real-time appearance-based person re-identification over multiple Kinect cameras," Int. Conf. Computer Vision Theory and Applications VISAPP, (2013).
[28] Wang, X., "Intelligent multi-camera video surveillance: A review," Pattern recognition letters 34, 3-19 (2013).