Demo: Real-time indoors people tracking in scalable camera networks

Vedran Jelača, Sebastian Grünwedel, Jorge Oswaldo Niño-Castañeda,
Peter Van Hese, Dimitri Van Cauwelaert, Peter Veelaert and Wilfried Philips
Department of Telecommunications and Information Processing, TELIN-IPI-IBBT,
Ghent University, Ghent, Belgium
E-mail: {vedran.jelaca, sebastian.gruenwedel, jorge.nino, peter.vanhese, philips}@telin.ugent.be

***Abstract*—In this demo we present a people tracker in indoor environments. The tracker executes in a network of smart cameras with overlapping views. Special attention is given to real-time processing by distribution of tasks between the cameras and the fusion server. Each camera performs tasks of processing the images and tracking of people in the image plane. Instead of camera images, only metadata (a bounding box per person) are sent from each camera to the fusion server. The metadata are used on the server side to estimate the position of each person in real-world coordinates. Although the tracker is designed to suit any indoor environment, in this demo the tracker's performance is presented in a meeting scenario, where occlusions of people by other people and/or furniture are significant and occur frequently. Multiple cameras insure views from multiple angles, which keeps tracking accurate even in cases of severe occlusions in some of the views.**

I. Introduction

Real-time people tracking is of fundamental importance in many vision applications, e.g. surveillance, elderly care or video conferencing. In real-world environments it is still a challenging task, mainly due to frequent occlusions of people and changes in the environments themselves (e.g. lighting changes). To handle occlusions multiple surveillance cameras are used to insure views on people from different angles. In classical (centralized) systems the cameras are just visual sensors connected to one or multiple servers with dedicated processing power, which limits the network expansion and creates communication load between the servers. To overcome the problems of centralized systems, in the recent years with development of smart cameras it is possible to distribute the computational load towards the cameras, creating scalable (decentralized) systems [1]. In this demo we show a decentralized scalable system working in an indoor setup.

II. System overview

Our system is constructed from multiple extrinsically calibrated smart cameras (simulated by IP cameras connected to dedicated PCs) and one fusion node (an additional PC) which estimates ground plane positions of people. The cameras perform tasks of foreground segmentation and 2-D tracking (tracking in the image plane), sending only the bounding boxes to the fusion node, as shown in Fig. 1a. The 2-D hypotheses are used in the fusion node to construct an evidence map for each person and from each camera view. These evidence maps are further fused into one final occupancy map for each person. For fusion of the evidences we use Dempster-Shafer (DS) theory [2], see Fig. 1b. The ground plane position for each person is estimated using a Bayesian filter approach and approximated as the ground cell with the highest probability. Furthermore, the ground plane positions with the predefined person's width and height are used to construct a person's cuboid, which is back-projected to each camera view and used to correct or verify the 2-D tracking hypotheses. 2-D tracking on the camera side is especially important when there is no frequent feedback from the server (e.g. due to processing or communication delays) since in those cases cameras themselves can still estimate positions of people and keep tracking them.

III. Demo description

The setup we use for this demonstration has one top view and four side view cameras. The cameras are placed on different sides of a room. The room is furnished with tables and chairs to resemble a meeting room. People are detected when entering the room and tracked within the room as illustrated in Fig. 2. We demonstrate the tracker's performance in a meeting scenario with the following events: entering the room, walking with different speeds (from standing still to walking fast), sudden changes of walking direction, people coming close to each other to shake hands, walking around tables to find a place to sit, moving chairs and tables to rearrange the room, sitting and standing still, changing a sitting position, bending over and under a table, and exiting the room.

On a camera side we also demonstrate the performance of our foreground segmentation method [3] based on detection of moving edges, which was developed to deal with sudden lighting changes. We also show the advantage of this method compared to two other foreground segmentation methods: Gaussian mixture model based, implemented by Zivkovic [4] and ViBe [5], see Fig. 3.

The scalability of the whole system is demonstrated by adding cameras to the system and showing that fusion processing remains real-time.

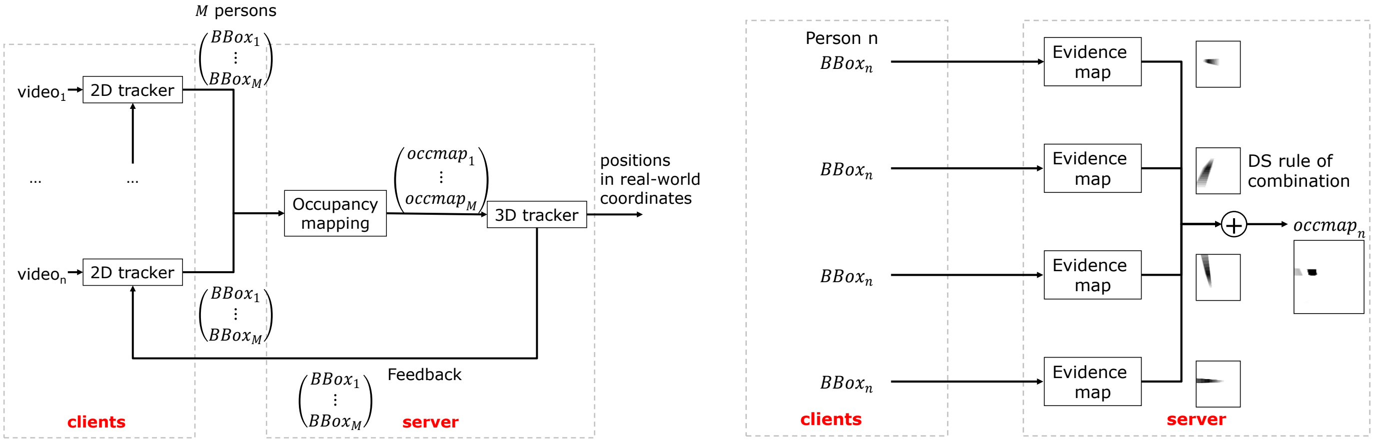

Fig. 1. The system architecture: a) Left: block diagram of the whole client (camera)- server (fusion node) architecture; b) Right: diagram of the occupancy mapping block.

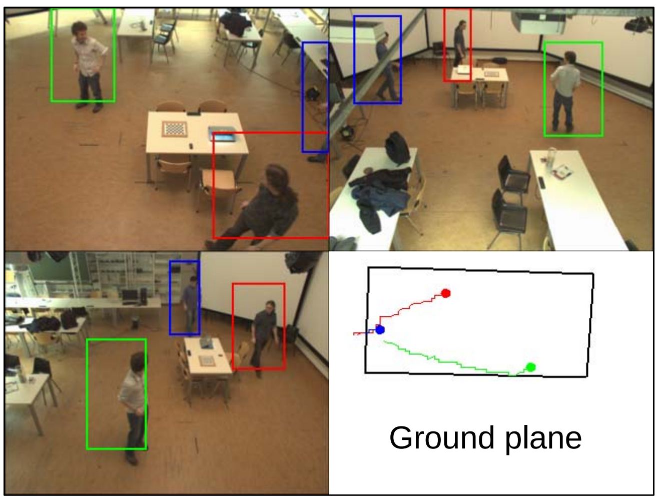

Fig. 2. Example of camera views and the tracking result. On a camera side 2-D hypotheses for different persons are represented with different colors. The same colors are used to represent the tracks and the current positions (visualized by colored dots) on the ground plane. The black rectangle on the ground plane image in the bottom-left represents the area around the table. The rectangle was used for a visual reference during the testing of the system.

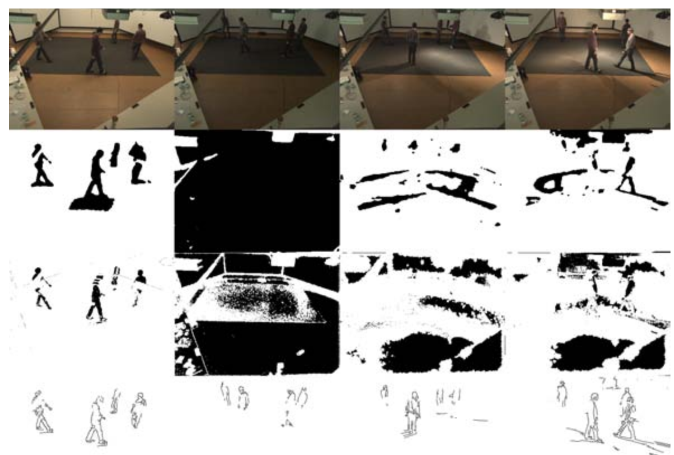

Fig. 3. Example of results of the compared foreground segmentation methods: first row original images, second row method of Zivkovic [4], third row ViBe method [5], last row our method based on detection of moving edges [3]. After the detection of moving edges we construct foreground blobs from these edges using convex hulls.

IV. Conclusion

In this demo we present a solution to overcome typical tracking problems of occlusions and non real-time performance, by using a network of smart cameras with overlapping views. This work is a good basis for further research towards creating real-time people tracker for even more complex and more crowded environments.

Acknowledgement

This work has been supported by the iCocoon project of the Flemish Interdisciplinary Institute for Broadband Technology (IBBT).

References

[1] M. Taj and A. Cavallaro, "Distributed and decentralized multi-camera tracking," *IEEE Signal Processing Magazine*, vol. 28, no. 3, May 2011.

[2] M. Morbee, L. Tessens, H. Aghajan, and W. Philips, "Dempster-Shafer based multi-view occupancy maps," *Electronics Letters*, vol. 46, no. 5, pp. 341–343, March 2010.

[3] S. Gruenwedel, P. Van Hese, and W. Philips, "An edge-based approach for robust foreground detection," in *Advanced Concepts for Intelligent Vision Systems*, 2011.

[4] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 2, no. 3, pp. 28–31, 2004.

[5] O. Barnich and M. Van Droogenbroeck, "Vibe: a powerful random technique to estimate the background in video sequences," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 945–948, May 2009.