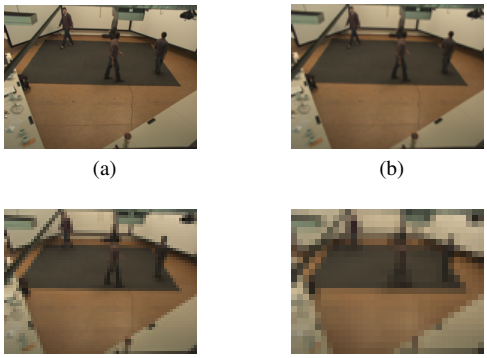
PhD Forum: Multi-view occupancy maps using a network of low resolution visual sensors

Sebastian Gruenwedel, Vedran Jelaca, Peter Van Hese, Richard Kleihorst and Wilfried Philips
Ghent University TELIN-IPI-IBBT
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
sebastian.gruenwedel@telin.ugent.be

***Abstract*—An occupancy map provides an abstract top view of a scene and can be used for many applications such as domotics, surveillance, elderly-care and video teleconferencing. Such maps can be accurately estimated from multiple camera views. However, using a network of regular high resolution cameras makes the system expensive, and quickly raises privacy concerns (e.g. in elderly homes). Furthermore, their power consumption makes battery operation difficult. A solution could be the use of a network of low resolution visual sensors, but their limited resolution could degrade the accuracy of the maps. In this paper we used simulations to determine the minimum required resolution needed for deriving accurate occupancy maps which were then used to track people. Multi-view occupancy maps were computed from foreground silhouettes derived via an analysis of moving edges. Ground occupancies computed from each view were fused in a Dempster-Shafer framework. Tracking was done via a Bayes filter using the occupancy map per time instance as measurement. We found that for a room of 8.8 by 9.2 m, 4 cameras with a resolution as low as 64 by 48 pixels was sufficient to estimate accurate occupancy maps and track up to 4 people. These findings indicate that it is possible to use low resolution visual sensors to build a cheap, power efficient and privacy-friendly system for occupancy monitoring.**

I. Introduction

Occupancy maps are an important step in many applications and are used for monitoring activities of people (for instance, how many people are in a room, the whereabouts of these people, etc.). Such maps can be accurately estimated using a distributed camera network over a single viewpoint setup.
However, next to the arising privacy issues, regular high-resolution cameras, which are usually used in such camera networks, make these systems expensive. Their high power consumption precludes battery usage, requiring more energy-efficient solutions. One possibility is the use of low resolution visual sensor networks (e.g. mouse sensors) [1], [2], but their limited resolution could degrade the accuracy of occupancy maps. It is also not clear which resolution is sufficient to construct accurate occupancy maps, that can be used for further processing.
In this paper, we simulate a visual sensor network to determine the minimal required resolution needed to construct these maps. To do so, we used a regular camera network and resized the image to simulate low resolution sensors (Fig. 1). In Section II we describe our data set which we used to perform simulations, followed by the architecture used to obtain the measures to determine the minimal resolution (Section III). Finally, we summarize our results in section IV.



(a) (b) (c) (d)

Fig. 1. Input images were resized for different resolutions (frame 530, camera 3): (a) 256x190, (b) 128x96, (c) 64x48 and (d) 32x24 pixels.

II. Data

For the simulation of low resolution visual sensors, we used a camera network in an 8.8m by 9.2m room. The dataset contains four people walking around the room observed by four cameras (780x580 pixels at 20 FPS) with overlapping views. Recordings were taken for about one minute during which ground truth positions of each person were annotated at one second intervals. These ground truth positions were used to measure the performance of our occupancy mapping and tracking for different image resolutions.

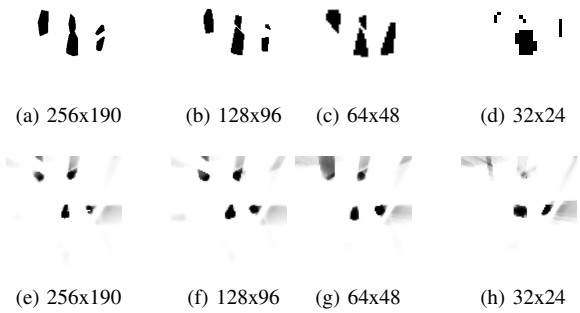
III. Methods

A. Foreground detection using moving edges

To perform foreground/background segmentation, we used a method to detect moving edges via analysis of the image gradient. The method uses edge dependencies as statistical features of foreground and background regions and defines foreground as regions containing moving edges. The background is described by a short- and long-term image gradient model using recursive smoothing for updating. The foreground mask (silhouettes of moving people) is obtained by clustering the moving edges and combining them via a convex hull technique (figure 2a-2d).

B. Dempster-Shafer based multi-view occupancy maps

The approach we followed (as described in [3]) constructs an occupancy map based on Dempster-Shafer reasoning [4],



(a) 256x190 (b) 128x96 (c) 64x48 (d) 32x24

(e) 256x190 (f) 128x96 (g) 64x48 (h) 32x24

Fig. 2. The upper row shows the foreground detection (foreground is indicated in black) for camera 3 and the lower one the corresponding occupancy maps (high evidence indicated in black) at frame 530. Compared to the others, the foreground detection performs poorly in the lowest resolution 32x24 (d) and so does the occupancy map of (h) wherein the evidence of one person is missing.

[5]. Such a map is calculated using different camera views and by fusing foreground silhouettes, obtained by the foreground detection III-A, onto a ground plane. This fusion of the evidences from all views uses the Dempster-Shafer's rule of combination (figure 2e-2h).

C. Multicamera tracking

Multicamera tracking was done using a similar method as the one described in [6]. The tracking method uses a Hidden Markov Model, which computes the target states using a Bayesian filter for state estimation under the assumption that the system can be modeled as a Markov process. Given the observation described by an occupancy map (III-B) at time instance t, the state transition matrix, and the initial probability distribution, the likelihood of an observation belonging to the state is computed.

D. Evaluation measures

1) p & n measures: As the first evaluation measure we use the n and p measures to evaluate the occupancy maps as described in [7]: n represents a measure of evidence at a person's position (within a radius of $10cm$, $n = 0$ is the ideal case) and p a measure of no evidence outside the positions ($p = 0$ is the ideal case). For p, we choose a radius of $70cm$ around the person's position. The ideal case for a method should be that $n = 0$ and $p = 0$, which means that all objects are detected and the evidence of a person is concentrated around the ground truth position.

2) tracking accuracy & tracking losses: The second evaluation measure contains the tracking results, namely the accuracy and the object losses. For that, we calculated the mean distance error to the ground truth positions followed by counting the number of losses for each person. A person is considered as lost if the distance between position and the ground truth is bigger than $70cm$. Those results give an indication about the accuracy of occupancy maps.

IV. Results & Discussion

To determine the minimal resolution, we performed an evaluation according to the measures in III-D. The results of

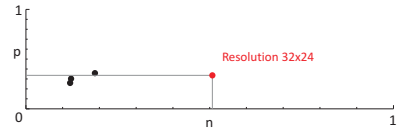


Fig. 3. The resolution 32x24 is out of range compared with the others that perform similar.

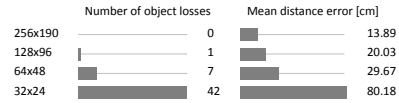


Fig. 4. The tracking result drops drastically for the resolution 32x24; for the other resolutions the decrease is acceptable.

the n and p measures (III-D1) are shown in figure 3. We see that there is a significant drop in accuracy of occupancy maps for the resolution 32x24 compared to the other resolutions (256x190, 128x96 and 64x48).

The tracking results (III-D2) prove this theses. In figure 4, the performance of the tracker using the lowest resolution 32x24 often fails due to inaccurate occupancy maps, meaning that the tracker produces inaccurate results ($80.18cm$ mean distance error from the ground truth) with many object losses.

V. Conclusion

We evaluated the dimension of a visual sensor for a room of 8.8m by 9.2m using 4 visual sensors. A resolution of 64 by 48 pixels is enough to build an acceptable occupancy map and to track four people. This implies that a visual sensor of around 64 by 48 pixels is fully sufficient to perform simple tasks. Hence, a mouse sensor could already be a cheap alternative to high-resolution cameras. Future work could include the robustness of a low resolution visual sensor network for tracking applications.

VI. Acknowledgment

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

References

[1] S. Hengstler and H. Aghajan, "A smart camera mote architecture for distributed intelligent surveillance," in *DSC'06*, Boulder, USA, 2006.

[2] S. Hengstler, D. Prashanth, S. Fong, and H. Aghajan, "MeshEye: a hybrid-resolution smart camera mote for applications in distributed intelligent surveillance," in *IPSN'07*, Cambridge, MA, USA, April 2007.

[3] 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.

[4] A. Dempster, "A generalization of bayesian inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 205–247, 1968.

[5] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, NJ, 1976, vol. 1.

[6] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 267–282, 2007.

[7] P. Van Hese, S. Grünwedel, J. Niño Castañeda, V. Jelaca, and W. Philips, "Evaluation of background/foreground segmentation methods for multi-view occupancy maps," in *Proceedings of the 2nd international conference on positioning and context-awareness (PoCA-2011)*, 2011, p. 37.