A LOW RESOLUTION MULTI-CAMERA SYSTEM FOR PERSON TRACKING

Mohamed Eldib[1,2], *Nyan Bo Bo*[1,2], *Francis Deboeverie*[1,2], *Jorge Nino*[1,2], *Junzhi Guan*[1,2], *Samuel Van de Velde*[1], *Heidi Steendam*[1], *Hamid Aghajan*[1,3] *and Wilfried Philips*[1,2]
[1]Department of Telecommunications and Information Processing, Ghent University-iMinds, Belgium
[2]Image Processing and Interpretation-Vision Systems
[3]Ambient Intelligence Research Lab, Stanford University, USA

ABSTRACT

The current multi-camera systems have not studied the problem of person tracking under low resolution constraints. In this paper, we propose a low resolution sensor network for person tracking. The network is composed of cameras with a resolution of 30x30 pixels. The multi-camera system is used to evaluate probability occupancy mapping and maximum likelihood trackers against ground truth collected by ultra-wideband (UWB) testbed. Performance evaluation is performed on two video sequences of 30 minutes. The experimental results show that maximum likelihood estimation based tracker outperforms the state-of-the-art on low resolution cameras.

Index Terms— Low resolution multi-camera systems, behavior analysis, tracking, foreground detection

1. INTRODUCTION

Multi-camera systems have recently emerged as a technology with multiple interesting applications such as improving, preventing and curing the wellness and health conditions of elderly, behavior analysis, tracking and surveillance scenarios. The current systems use high resolution cameras which increase the cost of installation. Although Passive Infrared Motion Sensor (PIR) is very cheap, and a popular choice among researchers, but still comes with disadvantages as inaccuracy, sensing motion (not presence) and relative position is important [14]. Also, Fabien et al. [3] state that a large number of PIR sensors could be required to cover most of the activities performed in a room (e.g. tracking). While, some images produced from a camera could pick up most of the activities. So, PIR still costly to maintain.

Low resolution processing is very difficult because the appearance of a person changes with the body movements, wide degree of variation in both pose and orientation, and variations in lighting. It is important to evaluate high resolution algorithms on low resolution cameras for modifications required to make them work better. PIR network could be simply replaced by a multi-camera system for the following: (1) PIR cannot sense people who are standing still [22]. In image, the detection of stand still person is possible, because the stand still person tend to move any parts of the upper body (head, shoulders and hands) and this could be easily detected by foreground detection algorithms, and (2) PIR output is highly bursty [22]. This limit PIRs system to single-person scenario. This is not the case of multi-camera systems, where tracking more than one person could be manageable.

In this paper, we propose a low resolution sensor network for the purpose of person tracking. We outline the main contributions of this paper as: (1) a low resolution multi-camera system and (2) detailed tracking comparison between probability occupancy mapping (POM) tracker of [2] and maximum likelihood tracker (ML) of [1] on low resolution cameras, against ultra-wideband (UWB) testbed.

The paper is organized as follows. In section 2, we present the related work. Section 3 gives an overview of the low resolution sensor network. Section 4 presents ultra-wideband (UWB) testbed. Section 5 describes our experimental results. Finally Section 6 draws conclusions.

2. RELATED WORK

There are several existing works on designing and building multi-camera systems. The authors of [4] construct a camera sensor networks for abnormal behavior detection in outdoors environment for short sequences (500 and 236 frames), and the resolution of the cameras is 320x240 pixels. Ian et al. [5] design an integrated mote for wireless sensor networks where the cameras are combination of medium resolution (CIF) and low resolution (30x30 pixels), and they demonstrated the use of their wireless sensor network with a single sensor node which produces frame of resolution 30x30, to count the pedestrian passing a walkway. Anthony et al. [6] present FireFly Mosaic, a wireless sensor network image system that has been deployed in apartment for activity analysis with image size 352x288 pixels. Instead of using low resolution cameras to analyze the occupancy map and to track people, Sebastian et al. [15] resized images captured by high resolution cameras.

Monitoring systems can be combined with different kind of sensors such as Passive Infrared Motion Sensor (PIR), microphone and magnetic switches to build hybrid systems. CASAS [7] monitors the activities of daily living to identify any different behavior patterns of dementia patients, while in Age In Place [8] aims at early illness detection of urinary tract infections for the elderly. On the

other hand, House_n project [9] provides tape-on sensors system for behavior analysis and monitoring.

Our approach of building multi-camera system is different from [4, 6, 15, 16], where we use only low resolution cameras, not high or medium resolution and not a combination of low and medium resolutions as in [5]. Our purpose is to reduce the cost of the sensor network, and to show how low resolution we can go to deliver good performance that matches the performance of high resolution cameras and the various employed sensors.

3. SYSTEM OVERVIEW

The system is composed of low resolution cameras. Each camera is controlled by a digital signal controller. The low cost comes from using ADNS3080 sensor [20] inside the cameras. This sensor is used in ordinary optical computer mouse. Camilli et al. [10] used this sensor with small adaptation to enable it to output video of 30x30 at over 100 frames per second. They opted for two image sensors to provide stereo information or near/far focusing with different lenses. The sensors connect over a Serial Peripheral Interface bus directly to the internal memory of the DSP which performs the insensor video processing. Camilli et al. [10] also opted for the dsPIC33EP512GP806, the most powerful controller available in the dsPIC family offering a maximum performance of 70 MIPS. It contains 536 kB of on-chip flash program memory and 52 kB of on-chip data RAM. An external 1 Mbit SRAM complements the internal memory. This allows video processing on cheap embedded micro-controllers rather than PCs, thereby reducing the overall system cost. Fig 1 shows the low resolution camera. For these reasons our multi-camera system uses the low resolution cameras of [10].

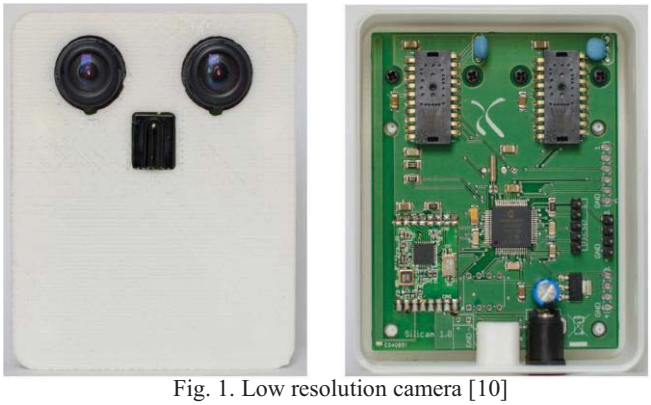


Fig. 1. Low resolution camera [10]

The low resolution sensor network is installed in a room of (7.5x4.7 m^2) but covering only (4.5x4.2 m^2). Fig 2 displays the room layout with camera positions. We calibrated the cameras using lighted sphere calibration method [11]. The system is divided into two main blocks. First, the captured images undergo texture-based foreground detection [12] (several methods were tested experimentally) which is proven to be robust to illumination changes, with updating the background model to remove the movements of non-human objects. Finally, tracker of Berclaz et al. [2] (POM) and maximum likelihood tracker of [1] (ML) use the detected foreground objects for person tracking.

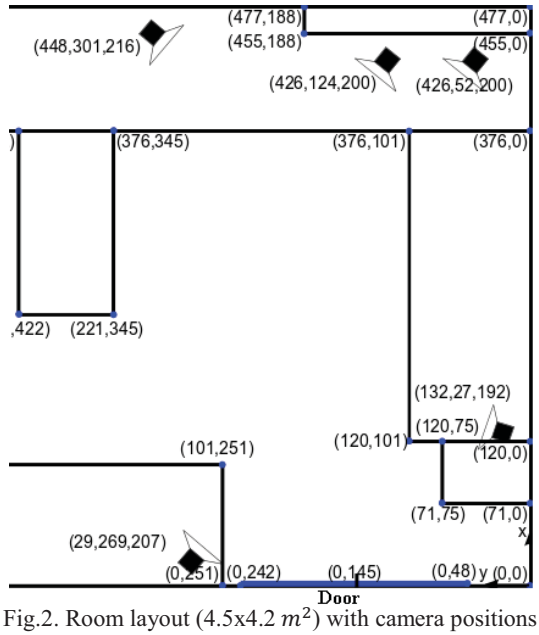


Fig.2. Room layout (4.5x4.2 m^2) with camera positions

3.1. Foreground detection

The images captured by the low resolution cameras suffer from noisiness, poor and quickly changing lighting conditions. We used ViBe [13] for foreground detection. The parameters of ViBe were tuned to obtain optimal detection. The obtained results were poor as shown in Fig 3(c). This shows that ViBe is performing badly under varying changes in illumination, and this is also reported by [12]. Next, we have adopted the correlation method because it is promising and robust to illumination changes [12]. First, the background model is built by computing the average of all frames, when there is no person present in the scene. Then, the correlation coefficient of pixels of the captured image and the corresponding pixels of the background model within a sliding window is computed by the following equation:

$$\rho(r) = \frac{\sum_{r' \in \omega(r)} I(r') I_{bg}(r')}{\sqrt{\sum_{r' \in \omega(r)} I(r')^2 \sum_{r' \in \omega(r)} I_{bg}(r')^2}}, \quad (1)$$

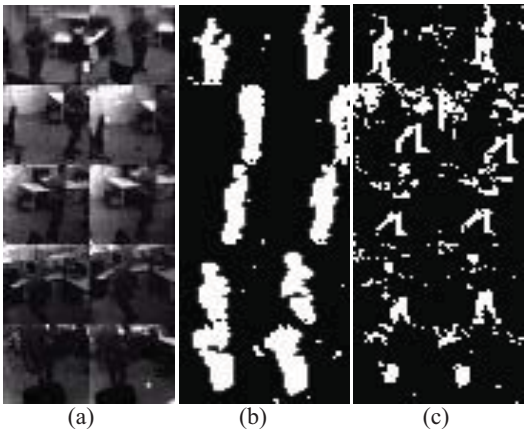
where $\omega(r)$ is a sliding window centered at r and $\rho(r)$ is the correlation coefficient between captured image pixel $I(r')$ and background model pixel $I_{bg}(r')$ over $\omega(r)$. Next, a pixel r is decided to be either foreground or background if:

$$F(r) = \begin{cases} 1 & \text{if } \rho(r) < \rho_{min} \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where ρ_{min} is the correlation threshold and varies between 0 and 1. The final result is a binary image F with white pixels representing the foreground object, and black pixels representing the background. The background model is updated according to learning rate α to remove false foreground detection of non-human objects such as chairs: $I_{bg}(r) = (1-\alpha)I_{bg}(r) + \alpha I(r)$. Fig 3 (b) shows the result of correlation method. We used window size of 10x10 as in [12], and it did not work well. Several window sizes have been experimented. Table 1 summarizes the tuned parameters to produce the best system performance.

Parameters	Values
Size of ω	2x2
ρ_{min}	0.98
α	0.005

Table.1 Correlation method tuned parameters



(a) (b) (c)

Fig. 3. (a) original, (b) detected foreground by correlation and (c) detected foreground by ViBe

3.2. Person tracking

We study the problem of person tracking on low resolution cameras with two trackers (1) maximum likelihood tracker of [1] and (2) tracker of Berclaz et al. [2].

Our maximum likelihood tracker of [1] relies on foreground detection and fusion center. At time t each camera computes the likelihood of the positions of a person within its viewing range based on the previous computations at time $t-1$ fed back by the fusion center. These likelihoods computed by cameras are sent to the fusion center in which the mostly likely position of a person at time t is computed. The final estimates are then fed back to all cameras with a single frame delay and are taking into account when estimating person's positions at time $t+1$.

The tracking system in [2] utilizes the concept of probabilistic occupancy mapping to find the person's positions. Then, the k-shortest path algorithm is applied to find the trajectories with the known positions of a person. This system needs an input of the whole video sequence or a batch of frames.

4. COMPARISON TO UWB GROUND TRUTH

In order to collect ground truth data for evaluating the performance of the tracking algorithms. We use the ultra wideband (UWB) testbed. It is important to note that UWB can only be used in lab experiments not in real life scenarios, because the maximum battery life is only 30 minutes and the person has to hold the mobile terminal all the time.

The UWB positioning data was collected using a testbed comprising 6 PulseOn P410 [21] UWB ranging devices, of which 5 were used as fixed anchors and 1 as the mobile terminal. The 5 fixed anchors were placed nearby the cameras positions. The PulseOn ranging devices used in the testbed are specifically designed to make range measurements with centimeter accuracy, by transmitting over a spectral bandwidth of 2.2GHz. In the experiment, the mobile terminal periodically collects range measurements from all anchors, and to enable real-time positioning, the low-complex linear least squares algorithm [19] was used for the estimation. We noticed that a small percentage position estimates were heavily corrupted when too many anchors were not in line-of-sight. So, we eliminated these sources of errors by discarding all inconsistent location estimates (i.e. a residue larger than 5 m^2[19]). In total, we captured two sequences of UWB position estimates, each time of one person for 30 minutes at 2.5GHz.

5. EXPERIMENTS

We used the total average tracking error (TATE) as in [18] to evaluate the performance of trackers. TATE is defined as the average of the Euclidean distances between positions estimated by the tracker and the corresponding UWB data positions in cm.

5.1. Results

Two videos of 30 minutes length were captured at 33 fps with their ground truth collected by the UWB data. The sequences contain only one person in the room performing walking and sitting activities. We use timestamp information from video sequences and UWB data to extract the trackers points, and the ground truth points within one frame difference. The UWB mobile terminal was not attached to the person. The device was approximately on 10 cm distance. So, the set of experiments are performed with the device displacement being taken into account. The generated tracks from trackers and ground truth are smoothed by applying the mean filtering.

We first measure the TATE and the total duration in seconds of good versus bad tracks for trackers. The

TATE results on two videos sequences are shown in Table 2. The TATE of our tracker [1] (ML) is 53.86 and 55.44 which is significantly smaller than the 111.50 and 73.00 of the state-of-the-art tracker [2] (POM). The total duration of good tracks, and bad tracks are similarly shown in Table 3 for both trackers. The ML [1] total duration of good tracks is 485 and 573 seconds which is noticeably higher than 242 and 330 seconds of POM [2]. The duration of a track is considered good if its TATE is less than 60 cm, otherwise it is a duration of bad track. Finally, Fig 4 and 5 show the number of good tracks at different threshold values, and it can be seen that our tracker has higher number of good tracks than Berclaz et al. [2] tracker at smaller threshold values. Fig 6 shows an example of tracks by trackers and ground truth. The green points indicate track from ground truth, and blue points indicate tracks from trackers. ML tracker produces more close tracks to ground truth than POM tracker.

The second set of experiment aimed at measuring the TATE of ML with correlation [12] and ViBe methods [13]. Experimental results are provided in Table 4 which shows the superiority of using correlation method [12] over ViBe [13] on low resolution cameras.

Sequence	ML(Ours)	POM [2]
1	53.86	111.50
2	55.44	73.00

Table.2 TATE (cm) comparison

Tracks	Seq 1		Seq 2	
	ML(Ours)	POM [2]	ML(Ours)	POM [2]
Good	485	242	573	330
Bad	13	256	120	363

Table.3 Total duration (sec) comparison of good and bad tracks

Sequence	ViBe [13]	Correlation [12]
1	118.24	60.037
2	79.94	59.15

Table.4 TATE (cm) comparison between ViBe and correlation

6. CONCLUSION

In this paper, we presented a low resolution sensor network as opposed to many systems that use medium and high resolution cameras, and we demonstrated with different foreground detection methods and trackers how good the performance can be under low resolution camera (30x30 pixels). We evaluated the performance of maximum likelihood tracker and state-of-the-art tracker against ground truth collected by UWB testbed on two video sequences. The experimental results show that the maximum likelihood tracker achieved high accuracy in person tracking compared to probability occupancy mapping based tracker. We compared two foreground detection methods ViBe and correlation. The evaluation shows that correlation is more robust to illumination changes on low resolution cameras.

7. ACKNOWLEDGMENT

This research has been performed in the context of the project “LittleSister” and the European AAL project “Sonopa”. This research has been financed by the agency for Innovation by Science and Technology (IWT), the Belgian National Fund for Scientific Research (FWO Flanders) and iMinds.

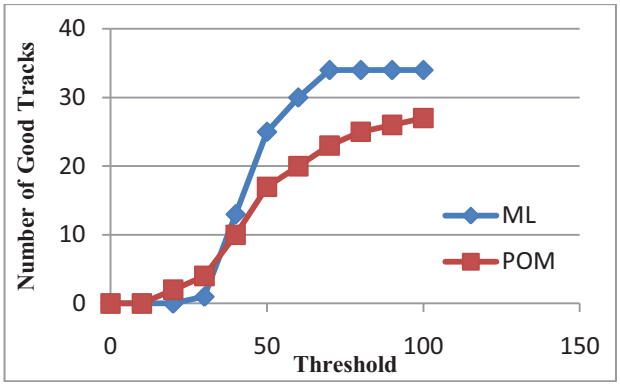


Fig. 4. Good track comparison between HT and FL on sequence 1

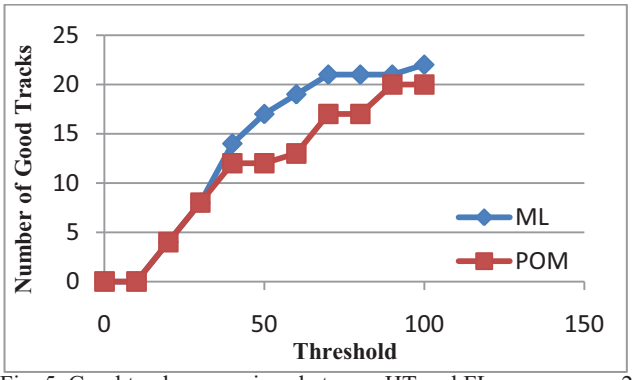


Fig. 5. Good track comparison between HT and FL on sequence 2

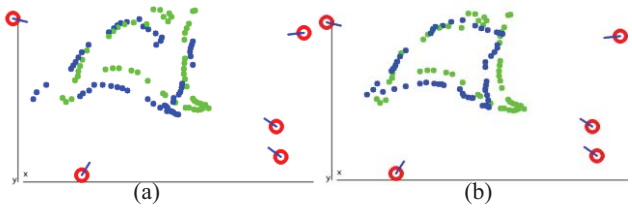


Fig. 6. Ground truth track (green) is plotted against (a) Barclaz et al. tracker [2] (blue) and (b) maximum likelihood tracker (blue)

8. REFERENCES

[1] B. Nyan, P. Van Hese, S. Gruenwedel, J. Guan, J. Niño-Castañeda, D. Van Haerenborgh, D. Van Cauwelaert, P. Veelaert and W. Philips, “Robust Multi-camera People Tracking Using Maximum Likelihood Estimation”. In Proc. Advanced Concepts for Intelligent Vision Systems (ACIVS) 2013, Poznan, Poland, pp. 584 - 595, 2013.

[2] J. Berclaz, F. Fleuret, E. Turetken and P. Fua, “Multiple Object Tracking Using K-Shortest Paths Optimization”. Pattern Analysis

and Machine Intelligence, IEEE Transactions on , vol.33, no.9, pp.1806,1819, Sept. 2011.
[3] F. Cardinaux, D. Bhowmik, C. Abhayaratne and M.S. Hawley, "Video Based Technology for Ambient Assisted Living: A review of the literature". In Journal of Ambient Intelligence and Smart Environments (JAISE). ISSN 1876-1364 (In Press), 2011.
[4] Y. Wang, D. Wang, and F. Chen, "Abnormal Behavior Detection Using Trajectory Analysis in Camera Sensor Networks". International Journal of Distributed Sensor Networks, vol. 2014, no. 839045, 2014.
[5] I. Downes, L.B. Rad and H. Aghajan, "Development of a mote for wireless image sensor networks". Proc. of COGnitive systems with Interactive Sensors (COGIS), Paris, France, 2006.
[6] A. Rowe, D. Goel and R. Rajkumar, "FireFly Mosaic: A Vision-Enabled Wireless Sensor Networking System". Real-Time Systems Symposium, 2007. RTSS 2007. 28th IEEE International , vol., no., pp.459,468, 3-6 Dec. 2007.
[7] P. Rashidi and D.J. Cook, "Keeping the Resident in the Loop: Adapting the Smart Home to the User". Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on , vol.39, no.5, pp.949,959, Sept. 2009.
[8] M.J. Rantz, M. Skubic, R.J. Koopman, L. Phillips, G.L. Alexander, S.J. Miller and R.D. Guevara, "Using sensor networks to detect urinary tract infections in older adults" e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on , vol., no., pp.142,149, 13-15 June 2011.
[9] E. Tapia, S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors". Lecture Notes in Computer Science, pp. 158-175, 2004.
[10] M. Camilli and R. Kleihorst, "Demo: Mouse sensor networks, the smart camera". Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on , vol., no., pp.1,3, 22-25 Aug. 2011.
[11] J. Guan, F. Deboeverie, M. Slembrouck and W. Philips, "Extrinsic calibration of multi-camera using sphere". To appear in ICIP 2014.
[12] B. Nyan, S. Gruenwedel, P. Van Hese, J.O. Niño-Castañeda, D. Van Haerenborgh, D. Van Cauwelaert, P. Veelaert and W. Philips, "PhD forum: Illumination-robust foreground detection for multi-camera occupancy mapping". Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on , vol., no., pp.1,2, Oct. 30 2012-Nov. 2 2012.
[13] O. Barnich and M. Van Droogenbroeck, "ViBE: A powerful random technique to estimate the background in video sequences". Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on , vol., no., pp.945,948, 19-24 April 2009.
[14] P. Rashidi, and A. Mihailidis, "A Survey on Ambient-Assisted Living Tools for Older Adults". Biomedical and Health Informatics, IEEE Journal of , vol.17, no.3, pp.579,590, May 2013.
[15] S. Grunwedel, V. Jelaca, P.V. Hese, R. Kleihorst and W. Philips, "Phd forum : Multiview occupancy maps using a network of low resolution visual sensors". In: 2011 Fifth ACM/IEEE Int. Conf. on Distributed Smart Cameras. IEEE (2011).
[16] H. Jiang, S. Fels and J. Little, "A Linear Programming Approach for Multiple Object Tracking". Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on , vol., no., pp.1,8, 17-22 June 2007.
[17] L. Zhang, Y. Li and R. Nevatia, "Global data association for multi-object tracking using network flows". Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on , vol., no., pp.1,8, 23-28 June 2008.
[18] S. Gruenwedel, V. Jelaca, J.O. Niño-Castañeda, P. Van Hese, D. Van Cauwelaert, P. Veelaert and W. Philips, "Decentralized tracking of humans using a camera network". Proceedings of SPIE, Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques, vol. 8301, 2012.
[19] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks: a quantitative comparison". Computer Networks, vol. 43, no. 4, pp. 499,518, 2003.
[20] ADNS-3080 data sheet, "High-Performance Optical Mouse Sensor".
[21] P410 data sheet. Time Domain Corp., Huntsville, AL, USA, http://www.timedomain.com/p410.php
[22] T. Teixeira, G. Dublon and A. Savvides, "A survey of human-sensing: Methods for detecting presence, count, location, track, and identity". In ACM Computing Surveys, Vol. V, No. N, 20YY, 2010.