Proceedings of the IEEE ICRA 2009
Workshop on People Detection and Tracking
Kobe, Japan, May 2009

Estimation of Pedestrian Distribution in Indoor Environments using Multiple Pedestrian Tracking

Muhammad Emaduddin
Robotics & AI Department
National University of Sciences and Technology
H-12, Islamabad, Pakistan
emaduddi@usc.edu

Dylan A. Shell
Computer Science Department
University of Southern California
Los Angeles, CA 90089, USA
dshell@robotics.usc.edu

Abstract* - We propose a two-tier data analysis approach for estimating distribution of pedestrian locations in an indoor space using multiple pedestrian detection and tracking. Multiple pedestrian detection uses laser measurement for sensing pedestrians in a heavily occluded environment which is usually the case with most indoor environments. . We adapt a particle filter based multiple pedestrian tracker to address the constraints of a limited number of sensors, heavy occlusion and real-time execution. Under these conditions any detection and tracking technique is likely to encounter a degree of error in cardinality and position of pedestrians. A completely new approach is employed which measures the error in tracker output due to occlusion and uses it to estimate a probability density function which represents the probable number of pedestrians located at a particular exhibit at a particular time. The end result of the system is a variable representing cardinality of pedestrians at a particular exhibit. This variable follows a distribution which is approximately normal where the variance of the probability distribution function is directly proportional to the error encountered by the tracker because of occlusion. The accuracy of our detection and tracking algorithm was tested both separately and in conjunction with the second-tier pedestrian distribution analysis and found marked improvement making our average pedestrian counting accuracy to at least 90% for all the pedestrian position data that we gathered with average pedestrian density at 0.34 pedestrians per sq. meter. Since the environment constraints for our system are unprecedented, we were unable to compare our result to any previous experiments. We recorded the number of people at each exhibit manually to establish the ground truth and compare our results.

I. Introduction

Indoor detection and tracking of pedestrians has a wide spectrum of applications ranging from architectural design of walkways to controlling pedestrian flow at public places like theatres, museums, airports, sports arenas, conventions centers and parks. Our effort in this paper is to devise a system capable of tracking and counting pedestrians in real-time using minimal resources. The word "minimal" here refers to the fewest possible laser measurement sensors with constraints on their orientation and placement. In real life applications, (e.g. narrow walkways, mounting on vehicles etc) the set of feasible locations for deploying sensors can be severely constrained. In our experience the requirements for non-intrusiveness of sensors i.e. reliable electrical power and maximum sensor coverage, limit the number and placement of sensors. Among the set of sensors that are available for tracking pedestrians, laser-range finders (LADAR) are presently among the most reliable and accurate; they reliably provide sub-centimeter accuracy at millisecond frequencies in range of environments. But even with the high fidelity that laser sensors provide, circumstances exist in which laser-based techniques fail to produce dependable pedestrian tracking results. While the techniques introduced in [1], [3], [4], [5], [6] and [7] are among the most successful in terms of tracking accuracy, they are significantly limited when dealing with occlusions [2] and many have a computational complexity that means they remain unsuitable for real-time applications. While our developed system is not as accurate as the online-learning tracker described in [4], it produces dependable results in heavily occluded environments while not compromising its real-time applications.

This work was partially supported by.US National Science Foundation under their Crosscutting Human and Social Dynamics (HSD) program.

II. Experiment Setup

Our test-bed for the detection and tracking algorithm consists of a tunnel like pathway which has five exhibits along its path and two access doorways to an unobserved theatre exhibit close to the centre of the pathway as shown in Figure. 1. Pedestrians can enter and exit the section of museum under discussion using any of the two accesses to the pathway. Pedestrians can also enter in and out from any of the doorway accesses to the unobserved exhibit. This pathway was chosen to be our test case as it allows various situations that can introduce complications in indoor pedestrian detection and tracking to be tested. These situations include: (i) Pedestrians move in a narrow tunnel like space thus there exists a high probability of occlusion due to close proximity of people: (ii) The pathway contains sections that can help us observe completely distinct behaviour of pedestrians e.g. at the exhibits where we expect pedestrians to stop and gather, away from exhibits where we expect pedestrians to walk with a relatively longer stride and at entrances where pedestrians are usually in an exploratory mode and tend to change walking direction very quickly: (iii) The two only access doorways to the circular theatre are observed by our laser scanners thus we were able to keep track of people present within the theatre without even directly observing them by simple count-keeping of people leaving and entering the theatre, (iv) Pedestrians visiting the exhibits were both adults and children which required us to tune detection to accept a relatively wide range of values for stride of a pedestrian, (v) Pedestrian groups,

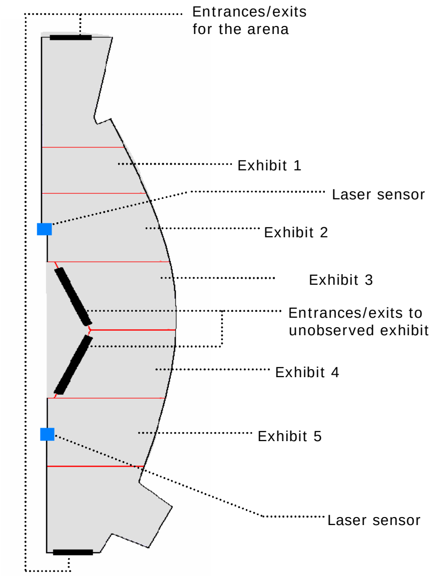


Fig.1. Test arena

which were usually a group of students lead by a teacher were a frequent occurrence at our test bed.

In order to meet our objective of tracking a fairly large number of people utilizing minimum possible resources, we decided to place two SICK Laser Measurement Sensors (LMS) 200 at a distance of approximately 8 meters from each other to cover an area of roughly 70 sq. meters. Ranges of our laser sensors overlapped for almost 16 sq. meters of area out of the total thus giving us a relatively accurate count in the overlapped area. The total area was divided into 5 cells each representing an exhibit (as shown with red lines in Figure 1). These cells will be later used to gather count of pedestrians visiting each exhibit at any given time. The off-the-ground height of rotating mirror within laser sensor was set at 29.9cm for all observations during the project. This height plays a crucial role in detection and association of clusters to the pedestrians since lowering the sensor height gives us discrete clusters representing feet but at the same time decreases our chances of detection of feet since we raise our feet while walking. On the other hand increase in height tends to ignore discrete clusters from feet of children or people with short heights. The effective scanning frequency of laser sensors is about 39Hz. The foreground points from the laser sensors were extracted easily by background learning and subtracting it from laser sensor readings.

III. The System

We present a system that is capable of detecting, tracking and the giving us the probability of pedestrian count at required locations. It comprises of two tiers explained in detail below

Tier 1: Detecting and Tracking Pedestrians

As will be shown, this involves a non-trivial adaptation and extension of the techniques developed in [1]. We describe the three parts below.

A. Clustering: Our algorithm starts by clustering incoming points from laser sensors using mean shift clustering algorithm. The system needs the size of cluster parameter at this point which is equivalent to the average area A of foot-print of an adult foot i.e. 0.04 sq. meters [10].

B. Temporal Correlation Analysis: After classification of points into clusters we iterate through clusters and establish which clusters belong to which pedestrian based on the notion that each pedestrian can be associated with a maximum of two clusters in *nth* frame which lie closest to the pedestrian in (*n-1)st* frame, we call this step as temporal correlation step. We divide this step into two phases (i) Phase one starts with identification of potential feet of pedestrians by calculating closest clusters and separating these as pairs. Only those clusters qualify as feet pair which lie within a parameter know as inter-feet distance *I* and have sizes in the vicinity of A sq. meters.

$$test\_pair(C_i, C_j) = \left\{ Pair^t(C_i, C_j) \middle| \begin{array}{l} distance(C_i, C_j) < I \\ \wedge \min\left(\sum_i \sum_j distance(C_i, C_j)\right) \\ \wedge \max(size(C_j), size(C_i)) \leq 0.04 \\ \wedge \neg Pair^{t-1}(C_i, C_j) \end{array} \right\} \quad (1)$$

The remaining unpaired clusters are thought to be clusters which are formed due to the fact that we cross our feet while walking thus rendering a single cluster in the laser sensor readings. The area of such clusters can be at most twice the footprint area of an average human foot (ii) Second phase consists of determining whether each cluster pair belongs to a newly detected pedestrian or it should be considered an update for an already tracked pedestrian *P* on the scene. This is done using association distance *D* that is the maximum distance that a pedestrian can travel between readings collected by laser sensors. Therefore the value of *D* is dependent upon the maximum walking speed of pedestrians in the arena.

$$associate\,(Pair_i^t, P_j^{t-1}) = \left\{ update\,(Pair_i^t, P_j^{t-1}) \middle| \begin{array}{l} distance\,(Pair_i^t, P_j^{t-1}) < D \\ \wedge \min(distance\,(Pair_i^t, P_j^{t-1})) \end{array} \right\} \quad (2)$$

Introducing above condition limits the distance travelled by pedestrians while being occluded and still being effectively tracked as a unique pedestrian.

We observed that the periodic motion of pedestrian feet described in [1] remains undetectable most of the time in environments cluttered with occlusions. Algorithm in [1] defines merge as a stage during walk when clusters of both feet of a pedestrian come close together and their clusters merge while split is described as a case when the pedestrian continues to walk after a merge and clusters of both feet split part. While merge and split cases were occasionally encountered during our experiment, we found out that detection of pedestrians in this manner is both inaccurate and computationally burdensome. The reason of inaccuracy lies in following notions (a) Most of the time we observe pedestrians walking in close proximity to other pedestrians or in the shape of groups, this tends to produce merges and splits that involve feet of two different pedestrians (b) Due to frequent occlusion (see Figure 2) We are likely to miss splits and merges belonging to a pedestrian thus rendering our split/merge detection mechanism useless under this situation (c) Pedestrians may not always walk, they might just stand for a while. Our solution to these problems as evident by (1) and (2) is to ignore the merge and split cases completely thus reducing the time complexity of temporal correlation step to $(n^2 logn + (nm).log(nm))/3$ where n is the number of clusters and m is the number of pedestrians on the scene. After this step, detected pedestrians along with their associated clusters are provided to the tracker i.e. our next step in sequence.

C. Tracking: The tracker is the component of our system that is responsible for estimating the parameters of motion and location attached with our pedestrian based on given updates from temporal correlation step. It uses a particle filter to estimate the position *p*, stride *s*, direction *d* and phase *ph* of a pedestrian as already employed in [1]. In brief the tracker keeps track of the pedestrians in three sub-steps (i) Update Step: Tracker weighs each pedestrian's particles proportional to their distance to the points belonging to its associated clusters: (ii) Sampling Step: After update step, the tracker randomly samples the weighted particles where the likelihood of any particle to be chosen is proportional to its weight. Thus a certain predefined number of particles *M* are chosen: (iii) Propagation Step: In the last step of tracking the sampled *M* particles are propagated through a multidimensional space representing the motion of the tracked pedestrian according to the walk model described in detail in [1]. This step modifies the position, stride, direction and the walking phase of a pedestrian and is performed without taking into account whether a pedestrian has received updates or not. The propagation of pedestrians that do not receive updates helps our tracker to track occluded pedestrians up till a certain amount of distance *D*.

During tracking each foreground point belonging to the pedestrian is used for calculating its distance with each of *M* particles belonging to the same pedestrian in tracker. For a maximum density of 1.8 pedestrians per sq. meter under which our tracker can perform optimally, it performs on the average nearly 504,000 calculations to update, sample and propagate 126 pedestrians through a single iteration. Given such high a penalty in terms of execution time, we deemed it extremely important for our algorithm to produce results with nearly same accuracy using fewer less computational resources in order to remain useful in real-time applications. Considering this requirement, we were able to successfully track pedestrians with very little degradation of accuracy by skipping unnecessary observations from laser sensors (See

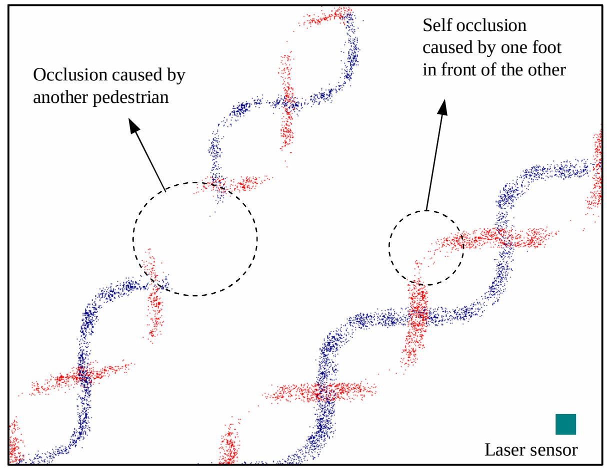


Fig. 2 An S-T representation of observable feet data

Table 1). The laser sensors provide our system with observations effectively after every 0.025 seconds. We forced our system to consider observations after every 0.05 seconds i.e. in effect dropping every second observation. This reduced the output accuracy by a very negligible value but the performance gain was more than 2 times. Since our system is specifically designed to handle occlusion, skipping an observation makes our system behave as if the skipped observation is due to an occlusion, thus by increasing the *D* parameter in temporal correlation module it compensates for most of the loss in accuracy.

The resultant system described up till now is relatively robust and accurate means of detecting and tracking pedestrians given the fact that we are performing these steps in real-time.

<u>Tier 2: Pedestrian Distribution Analysis</u>

Although reliability in the results could have been achieved by integrating techniques like online-supervised learning [4], Multiple Hypothesis Tracking [3] or Auxiliary Particle Filter switching [2] in the first tier, but doing so will exclude our tracker completely from the realm of real-time systems. Thus, the second tier of our system is designed to further enhance the reliability of the pedestrian count output for each exhibit while keeping the computational complexity growth nearly constant. We term this tier as the pedestrian distribution analysis tier as it is concerned with keeping track of pedestrians crossing in and out of each cell cells within the environment. A cell comprises of area in front of an exhibit defined using cell boundaries (as marked in Figure 1). By maintaining information about the distribution of people over cells, although the system cannot answer questions about where particular pedestrians are, one may still investigate questions about the flow of people and how their (average) route selection depends on the (average) presence or absence of people.

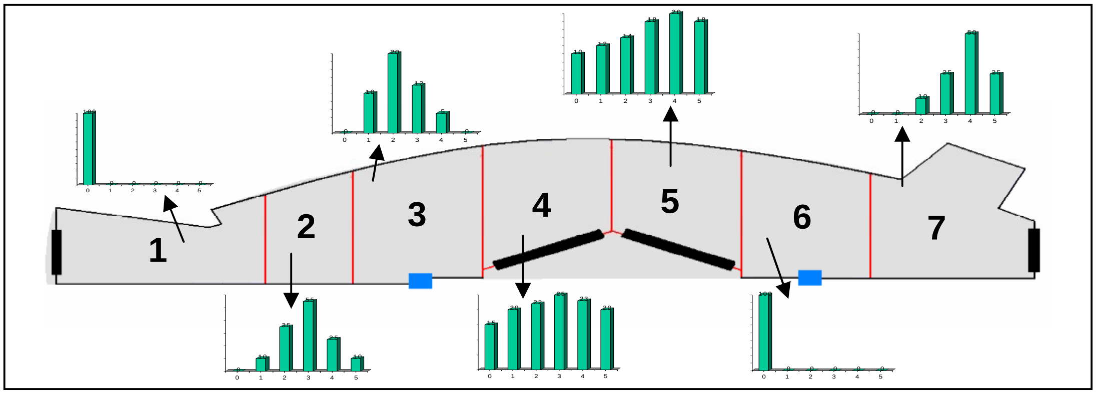


Fig.3. Pedestrian Distribution Analysis tier Output

Detecting number of people crossing into and out of each cell we were able to deduce the number of people N_i^t in each cell i at each time-step t. This number contains a certain error directly proportional to the percentage of the cell boundary hidden from laser sensors due to the pedestrians standing/walking very close to the laser sensors. In order to factor-in the error present in this number, we choose to represent the output of the system for each cell as a distribution over the number of people. A distribution variable X_i^t for each cell i at any given time t is a state of our belief that represents all past observations including the current one. This is achieved via updating the distribution variable X_i^t for each exhibit at each time-step. Variable X_i^t is defined as

$$X_i^t = \left\{ \pi_u^t, \quad u = N_i^{t-1} - r, N_i^{t-1} - (r-1), ..., N_i^{t-1} + r \right\} \qquad (3)$$

Here u is an index that runs through the range of weights π which represent our probability density function *(pdf)*. Most generally the range adjustment value r is subject to the requirement of the analyst which differs with the application of our system. (We used the physical capacity of the exhibits to place limits on this range of values.) Changing the value of r increases or decreases the domain of our distribution function. N_i^{t-1} is a number that has the maximum weight π_u^{t-1} associated to it in the distribution X_i^{t-1} from previous time-step. Following steps update the variable X_i^t at each time-step via a Gaussian update U_i^t whose variance is determined by the percentage of cell boundary occluded at any given moment. The update step is given below.

$$X_i^{t+1} = X_i^t + \delta^{t+1}(U_i^{t+1} - X_i^t)$$

$$\text{where} \qquad U_i^t = \left\{ \pi_u^t, u = N_i^{t-1} - r, ..., N_i^{t-1} + r \right\} \qquad (4)$$

$$\text{and} \qquad \delta^{t+1} = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_{t+1}^2}$$

Here if U_i^{t+1} has high variance relative to U_i^t then δ^{t+1} is small thus it has little impact on value of X_i^{t+1}. This ensures that updates which have more chance of error are factored-in less into our current belief X_i^{t+1}. σ_t^2 is the adjusted-variance in update distribution U_i^t and is determined using this intuitive criteria :

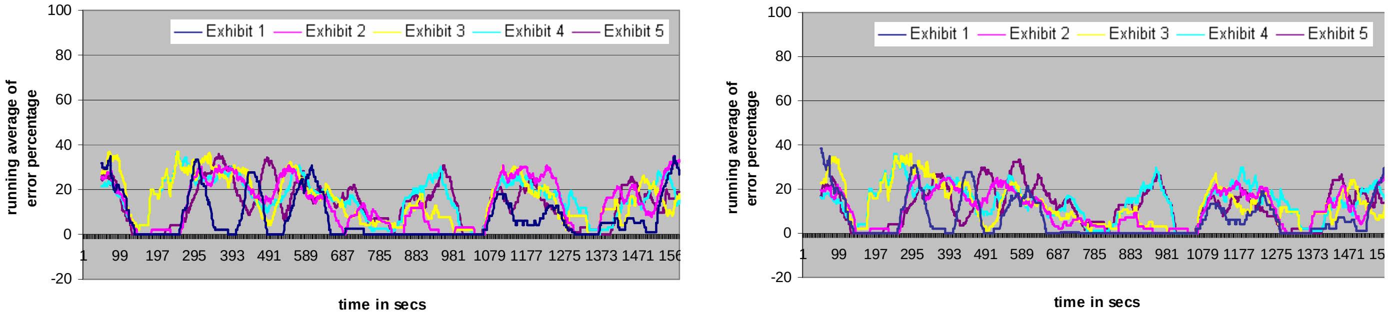
$$\sigma_t^2 \quad \infty \quad (g\,\sigma_t^2)(\frac{o_i^t}{l_i}) \qquad (5)$$

Here $g\sigma_t^2$ is the Gaussian variance of update U_i^t. The criteria described in (5), sets the variance to be directly proportional to the ratio of length of occluded boundary of cell o_i^t (calculated at every time-step) to the total visible length l_i of the cell boundary.

Pedestrian Distribution analysis tier thus represents snapshots of *pdfs* for each cell at each time-step which gives us a measured idea about the confidence that we can place on the pedestrian count in each cell (see Figure 3).

IV. Discussion

Tracking pedestrians at exits and entrances proved to be one of the trickiest parts during the system design. We know that the tracker output grows accurate with increase in the time for which a target is observed since tracker gets more chances to update and propagate its particles so that these can match target dynamics. Thus the places the tracker tends to be most inaccurate are the entrances to the observed area where the observed time for entering targets is limited. In order to estimate by what margin our tracker fails to track entering pedestrians, we performed an experiment by first measuring the number of pedestrians crossing east to west across a line dividing the observed area into two halves. We did this because our tracker is relatively accurate about pedestrians in the middle of observed area since the tracker had enough time to track these pedestrians. Then we considered the same line as an entrance and ran the tracker for the second time on the same



(a) Error before Tier-2 application

(b) Error after Tier-2 application

Fig. 4 System counting error comparison

set of observations for people entering in east to west direction considering updates only from one half of the observed area and ignoring the rest. The difference between the numbers of people crossing east to west in both cases provided us with the bias the tracker had in tracking pedestrians near the entrances. We used this bias b_i in following manner to adjust the number of people in cells that are situated at the entrances:

$$N_i^t = N_i^t + b_i$$

Using updated cardinality as an input to the second tier of our system proved to be beneficial in terms of accuracy but we restrained to declare it a formal part of our system since it would make tedious experimentation to learn bias, a prerequisite for deploying our system thus limiting its applications.

V. Results

We tested our system in terms of accuracy and computational efficiency. In data collection phase we manually recorded the pedestrian crossings over certain episodes of time observed via laser data stream for each of the cells. These time-stamped recordings were accurate up to 1 second resolution and served as our ground truth. For accuracy measurement we computed following two errors. (i) ($N_i^t - ground\_truth_i^t$) for exhibits *i=1 to 7* (Figure 4a shows a single episode depicting the error for each of the cells). Here error is calculated using pre tier-2 measurement i.e. N_i^t from tier-1. Here the cumulative average counting error for all our observations for all the exhibits totalled to be 13.8%. (ii) ($\mu_i^t - ground\_truth_i^t$) for exhibits *i=1 to 7* where μ_i^t is the value with highest probability in the *pdf* representing X_i^t (Figure 4b shows the same episode as shown in fig. 4a depicting the error for each of the cells). This error is computed using output from tier-2 of our system. The average counting error for all our observations for all the exhibits in this case stood at 9.83% which shows marked improvement as a result of applying tier-2.

By applying our tier approach to laser data collected by recording over 50 hours of museum visitors, we are able to plot locations of high-traffic. This is shown in Figure 5 using a colour coded scheme in which red highlights reflect the positions that people spend most of their time in. In a sense, this represents the time-averaged distribution from tier-2.

VI. Conclusion

Techniques described in [1], [3], [4] and [6] stress the tracking accuracy. Our effort is focused on retrieving analysable results using fast tracking techniques in order to get reliable pedestrian count in heavily occluded environments. Our pedestrian detection and tracking algorithm is extremely computationally intensive as is the case with all other multiple target tracking algorithms [7] and this happens in our case due to computations like inter-cluster, cluster to pedestrian distance calculation and propagation of a high number of particles in particle filter at each time-step. During our experiment phase we were able to produce sufficiently accurate results in a more reliable format for scientific analysis of pedestrian distribution in indoor environments.

Acknowledgements

Support from Interaction lab, University of Southern California (USC) is gratefully acknowledged. Also support from all undergraduate students who worked under NSF's Research Experience for Undergraduates (REU) program is appreciated. Scholarship grant from Fulbright Commission is acknowledged and appreciated as it funded the research assistantship for one the authors. We thank Professor Kristina Lerman from USC Information Sciences Institute for her constant advice and mentoring during all phases of our research. Lastly this work was made possible by the motivation given to us by our ever helpful Professor Maja Mataric.

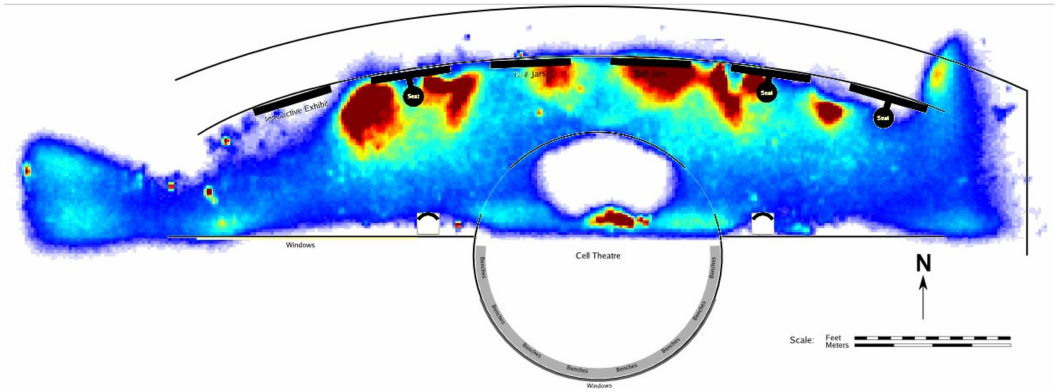


Figure 5: Locations of high traffic within the museum exhibit

TABLE I
COMPUTATIONAL EFFICIENCY FOR VARYING PEDESTRIAN DENSITY (System: Ubuntu 8.04, kernel 2.6.24-18, Intel Pentium Mobile 1700 MHz Processor)

Frame skip rate	Average execution time for 1 sec of frames	Peak density encountered (people per sq. m)	Average density (people per sq. m)	Average counting error % (error/truth*100)
Every 2 out of 3	0.58 sec	0.33	0.10	11.8
Every 2 out of 3	0.8 sec	1.94	0.35	11.9
Every 2 out of 3	0.92 sec	0.72	0.54	13.6
Every other	0.71 sec	0.33	0.10	9.7
Every other	0.94 sec	1.94	0.35	9.4
Every other	1.07 sec	0.72	0.54	10.2
None skipped	1.94 sec	0.33	0.10	8.5
None skipped	2.6 sec	1.94	0.35	8.4
None skipped	3.12 sec	0.72	0.54	9.3

REFERENCES

[1] Shao .X, Zhao .H, Nakamura .K, Katabira .K, Shibasaki .R, "Detection and Tracking of Multiple Pedestrians by Using Laser Range Scanners" in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, April 2007.

[2] Bando .T, Shibata. T, Doya. K, Ishii. S, "Switching Particle Filters for Efficient Real-time Visual Tracking" in *Proceedings of the 17th International Conference on Pattern Recognition 2004*, vol. 2, pp. 720-723, Aug 2004.

[3] Arras .K, Grzonka .S, Luber .M, Burgard .W, "Efficient People Tracking in Laser Range Data using a Multi-Hypothesis Leg-Tracker with Adaptive Occlusion Probabilities" in *2008 IEEE International Conference on Robotics and Automation*, pp. 1710-1715, May 2008.

[4] Song .X, Cui .J, Wang .X, Zhao .H, Zha .H, "Tracking Interacting Targets with Laser Scanner via On-line Supervised Learning" in *2008 IEEE International Conference on Robotics and Automation*, pp. 2271-2276, May 2008.

[5] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, pp. 843–854, Dec 1979.

[6] Wang .J, Makihara .Y, Yagi .Y, "Human Tracking and Segmentation Supported by Silhouette-based Gait Recognition" in *2008 IEEE International Conference on Robotics and Automation*, pp. 1698-1703, May 2008.

[7] Khan .Z, Balch .T, Dellaert .F, "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, Issue. 11, pp. 1805-1819, Nov. 2005.

[8] Thrun, S., "Particle filters in robotics", *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*, 2002.

[9] Hollinger .G, Djugash .J, Singh .S, "Tracking a Moving Target in Cluttered Environments with Ranging Radios", in *2008 IEEE International Conference on Robotics and Automation*, pp. 1430-1435, May 2008.

[10] Hawes, Michael R., "Quantitative morphology of the human foot in a North American population" in *Ergonomics*, Vol. 37, Issue. 7, pp 1213, 1994.