Proceedings of the 2006 IEEE International Conference on Robotics and Automation
Orlando, Florida - May 2006

People tracking and following with mobile robot using an omnidirectional camera and a laser

Marin Kobilarov and Gaurav Sukhatme
Robotic Embedded Systems Laboratory
University of Southern California
Los Angeles, CA, USA

Jeff Hyams and Parag Batavia
Applied Perception
Cranberry Township, PA, USA

Abstract— **The paper presents two different methods for mobile robot tracking and following of a fast-moving person in outdoor unstructured and possibly dynamic environment. The robot is equipped with laser range-finder and omnidirectional camera. The first method is based on visual tracking only and while it works well at slow speeds and controlled conditions, its performance quickly degrades as conditions become more difficult. The second method which uses the laser and the camera in conjunction for tracking performs well in dynamic and cluttered outdoor environments as long as the target occlusions and losses are temporary. Experimental results and analysis are presented for the second approach.**

I. INTRODUCTION

We describe a mobile robotic system for tracking and following moving people. Such a system provides important capabilities for human robot interaction and assistance of humans in various settings. That is why tracking with mobile robots has been an active research area with many successful systems developed, such as for museum guidance [1], hospital assistance [2], or pedestrian tracking [3]. Most of these systems are designed to operate indoors or in flat and structured outdoor environments and at limited robot speed. Our goal, on the other hand, is to enable a mobile robot to track and follow a fast moving person (of speeds near 3 m/s), in an outdoor, potentially cluttered, and dynamic environment where the sensors are subject to more noise and the tracked person can be occluded or lost for short periods of time. Such a capability is important for robots assisting humans in rougher and more difficult environments. Typical applications include disaster-rescue missions or military mules carrying equipment and assisting soldiers.

Tracking under the above requirements is harder and often systems proved to work well normally are ineffective in such adverse conditions. Thus, we think that it is useful to describe our experience building two such systems, each having its own benefits and shortcomings. The first system we describe detects and tracks motion from visual input only and uses a laser range-finder to estimate the range to the centroid of motion in the camera image, while the second approach uses the laser as primary sensor and camera as secondary.

The first approach uses egomotion compensation in transformed perspective images and frame differencing to detect motion. A particle filter is then used to track the motion and an EM algorithm to cluster and find its centroid [4]. The second system uses the laser to extract the 3D relative position of blobs that might have originated from a person and uses these as measurements to a probabilistic data association filter (PDAF). The PDAF is augmented using "feature likelihoods" for each measurement that correspond to a color distribution template match of an area in the image covering the corresponding laser blob. The main contribution of this paper is to present the extension and analysis of existing methods in a more difficult setting than previously tried by others.

Our platform is the two-wheel dynamically balancing Segway Robot Mobile Platform (RMP) which is well suited for outdoor conditions. It has maximum speed of 3.5 m/s, can climb slopes of 17^o (without load) and with a good set of tires can overcome small obstacles and uneven terrain. The robot has a human-size footprint and can carry loads up to 100 lbs. It is equipped with a passive SICK laser and an omnidirectional camera. While the omnicam has a 360^o field of view and a single center of projection, a single camera alone cannot provide reliable range to the tracked person, since it is usually moving along the line of robot translation (thus accurate triangulation from multiple views is not possible). On the other hand, the laser provides range but only in a single plane and we do not assume that the laser sensing plane always intersects with the person (e.g. the robot could be pitching too much, or the human might simply be on a different terrain elevation out of laser sight). If tracking and following are stable enough though, the robot can be guaranteed to always be close to the person therefore the terrain between robot and person can be approximated by a plane. Alternatively, it is possible to use stereo vision but a major requirement for this project is a very wide field of view in front of the robot in order to guarantee stable tracking of a fast maneuvering person. At the time this project was developed we did not consider an actively controlled stereo head or a wider field of view (i.e. fish-eye) stereo pair. These options remain to be explored in future work.

The main requirements for the system include reliable motion detection, motion tracking through accurate estimation of the state of the moving person and stable following of the person.

0-7803-9505-0/06/$20.00 ©2006 IEEE

II. Related Work

There are a variety of human visual tracking approaches that differ mainly in the feature extraction and motion modeling. Aggarwal and Cai [5] provide an overview of interpreting and tracking human motions using cameras. Most methods extract and track features based on simple cues such as color, texture, or shape contour. There are a number of different filtering techniques to estimate the targets states. Rasmussen and Hager [6] use a joint probabilistic data association filter (JPDAF) to track multiple parts of the human body by extending the filter to include constraints between the parts. MacCormick and Blake [7] use a sampling based joint data association filter for multiple object tracking based on probabilistic exclusion. Cox and Hingorani [8] use multiple hypothesis tracking (MHT) to resolve association uncertainties between corner feature tracking using Kalman filters. Davis et. al. [3] use the condensation algorithm with quasi-random sampling to track a deformable shape model of humans. Jung and Sukhatme [4] use egomotion compensation, and track the difference image between two frames using a particle filter.

Motion model based tracking using range and bearing (e.g. from laser range-finders) has been an active area of research for decades, mostly inspired by applications for air and ground vehicle and missile tracking using radar or sonar sensors [9], [10]. Such systems employ various filtering techniques such as Kalman filter (with extensions such as cascading or interactive multiple model filters(IMM)), PDAF, JPDAF, MHT, set theory based filtering. Such techniques can be directly applied to tracking humans using a mobile robot. One system which extends the JPDAF method is the sampling-based JPDAF developed by Schulz et. al. [11] that replace the underlying EKF in the JPDAF with a particle filter. They also employ an occlusion probability map to explicitly deal with target occlusions in the filter.

There have been several research groups which have also successfully used the Segway RMP for tasks involving tracking and following of moving people (e.g. [12]). These works use simplified and less robust tracking methods to demonstrate higher level capabilities.

III. Sensor configuration

The robot (fig. 2) is a rigid body operating in 3D workspace. Let the robot frame correspond to a transformation ${}^{A}\mathcal{T}$ of a fixed global frame. The laser and camera are rigidly attached to the robot with corresponding transformations (with respect to the robot frame) ${}^{L}_{A}\mathcal{T}$ and ${}^{C}_{A}\mathcal{T}$. A laser range-finder reading of range d and bearing β corresponds to a 3D point in homogeneous coordinates:

$$P = {}^{A}\mathcal{T}\,{}^{L}_{A}\mathcal{T} \begin{pmatrix} d\cos\beta \\ d\sin\beta \\ 0 \\ 1 \end{pmatrix}$$

The same point in the camera coordinate frame is

$${}^{C}P = \begin{pmatrix} {}^{c}x \\ {}^{c}y \\ {}^{c}z \\ 1 \end{pmatrix} = ({}^{C}_{A}\mathcal{T}^{-1})({}^{A}\mathcal{T}^{-1})P$$

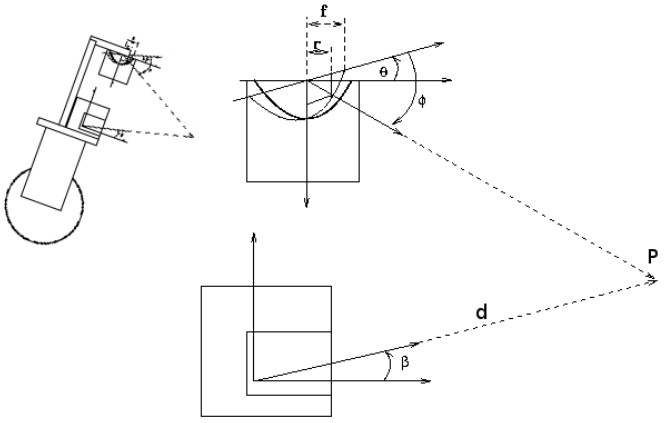

Fig. 1. Laser and omnidirectional camera configurations

.

We have experimented with two catadioptric adapters: a NetVision360 by Remote Reality, and a larger model made by EyeSee360. The two types have similar configurations and here we briefly describe the NetVision360 setup. It consists of a parabolic mirror and orthographic lens. The camera's intrinsic parameters include only the raw image center (x_c, y_c) and the combined projection focal length f. These three parameters are easily determined by direct observation of the image since the lens mirror is entirely visible. We assume that the lens mirror axis is parallel to the camera axis and we do not account for skew of the image sensor. Alternatively, the camera could be automatically calibrated using the parallel line method [13] or [14] using feature correspondences. Using the mirror transformation derived in [15] point P projects onto image coordinates (u, v) according to:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x_c + r\cos\theta \\ y_c + r\sin\theta \end{pmatrix},$$

$$r = f(\sqrt{\tan^2\phi + 1} - \tan\phi)$$

$$\theta = atan2({}^{c}y, {}^{c}x), \phi = atan2({}^{c}z, \sqrt{{}^{c}x^2 + {}^{c}y^2})$$

Fig. 2. Segway RMP

The spherical coordinates θ and ϕ are the azimuth and elevation of ${}^{C}P$ and r is the radial coordinate of its projection in the camera frame (fig. 1). The inverse transformation from image coordinates to a unit vector pointing at point P in the global frame can be computed accordingly. The equations above define the geometric relationship between the measurements of the two sensors.

The robot pose ${}^{A}\mathcal{T}$ is computed from the on-board gyros providing pitch and roll and from odometry providing yaw and position. Although we have implemented a fast full pose egomotion estimation (visual odometry) algorithm using the panoramic camera based on Bruce and Horn's method [16], our recursive tracking methods are only sensitive to the frame-to-frame pose estimation error which justifies using unfiltered odometry and gyro readings.

IV. Egomotion compensation and frame differencing method

Jung and Sukhatme [4] show that frame differencing can be used to detect and localize motion in sequences of monocular perspective images. They track corresponding features in each frame using the Lukas-Tomasi-Kanade (LTK) feature tracker. They compensate for egomotion by computing a bilinear pixel transformation between consecutive frames using linear regression on the tracked correspondences. The aligned images are then differenced and the resulting grey-scale image is used as a measurement to a particle filter. In order to use this system we transform the raw omnidirectional image into a set of perspective images with equally spaced (i.e. 45 degrees) view axes and the same field of view. A pixel in the omnidirectional image is transformed to a pixel in a perspective image by first finding the vector parallel to the corresponding world point in the camera frame (see sec. III) and then finding the intersection of this vector and a perspective image plane. In order to avoid the effect of aliasing, optic flow (using LTK) is performed on the original raw image and the coordinates of each pair of corresponding features are transformed to local coordinates in the perspective image they project onto. Then egomotion compensation and frame differencing is done separately on each perspective image, and the combined result is again input into a particle filter. An EM clustering algorithm is used to localize the peaks of the resulting particle motion distribution.

A benefit of this approach is that motion is tracked only from visual input and the laser is only used to find the range of each moving object. If the object is not likely to intersect the laser sensing plane then tracking can still continue, for example, by approximating the range using prior knowledge. A major shortcoming of this method though is that the egomotion compensation step assumes that camera images originate from an affine projection. In the setting of our project, a scene can be cluttered with close obstacles, such as trees, corners, or other people. Such conditions violate the affine transformation and the output from frame differencing becomes unreliable. While this is not so problematic during slower motion in obstacle free environment, the performance of the system quickly degrades at higher speeds . For example, fig. 3 shows a short segment

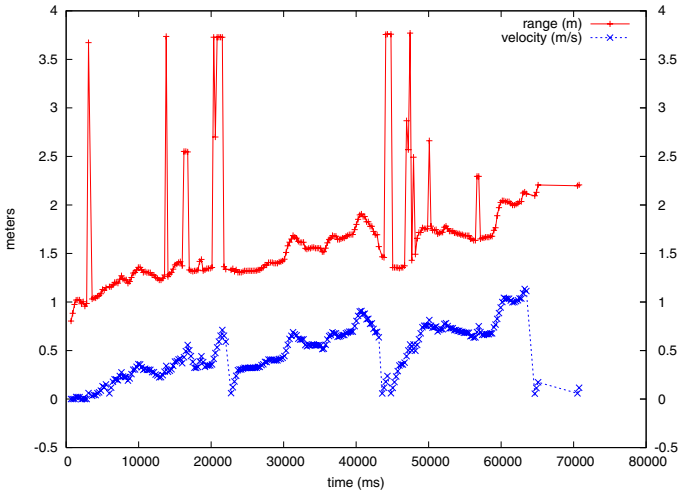

Fig. 3. Typical run of the system: target losses occur frequently and above 1 m/s tracking becomes unreliable

of a typical run among obstacles where after occasional losses during slow speed, at the end of the run, after the robot has reached 1 m/s the target is lost.

Although the method was not suitable for high-speed tracking among obstacles, we presented it since we think it is beneficial to describe how adverse conditions affect methods that normally work well. Such problems are likely to arise in similar motion tracking methods based on monocular vision.

V. Laser and Camera augmented PDAF Tracking

While the previous system tracks moving objects from visual input only, the second system we implemented uses the laser as the primary sensor. Each laser scan is processed into connected components (blobs) that might have originated from people in the laser field of view. Regions in the image corresponding to each of these objects are then used as additional measurements. Thus, multiple hypothesis are formed about the position and appearance of the tracked target.

We use a probabilistic data association filter (PDAF) which is an all-neighbor approach to data association under the assumption of a single target in clutter. It is proved to be more robust than a nearest neighbor data association scheme (i.e. single measurement Kalman filter) since its correction step uses a mixture of all observations residuals and also accounts for the possibility that none of the observations are valid [10]. We have implemented two target maneuver models: constant velocity, white noise acceleration model (using Cartesian coordinates and their velocities in the global frame as the target state), and a nonlinear coordinated turn model with a state vector containing the target turn rate in addition to the constant velocity model state components. While both models can capture the random dynamics of a fast moving person, the second model performs better in cases of target occlusion since it can account for nonlinear trajectories and predict more accurately where the target might reappear after occlusion. We apply two levels of gating: a maneuver gate based on the maximum possible velocity and acceleration of

a moving person, and the standard elliptical covariance-based gate. The PDAF, as well as the maneuver models used are thoroughly described in [10].

We treat the visual data as feature measurements and augment the PDAF by computing target-to-clutter feature likelihoods [9]. A kinematic measurement $\mathbf{y}_j$ is the position P_j of the center of laser blob j in a global frame. Since we assume that this point belongs to a person then we can "lift" it to an appropriate height h (i.e. centered at the person's chest) that would correspond to the center of a visual region of interest (ROI). Let the x-y global coordinates of the edges of the blob be (x_j^l, y_j^l) and (x_j^r, y_j^r). Let camera spherical coordinates (θ_l, ϕ_l) and (θ_r, ϕ_r) correspond to points (x_j^l, y_j^l, h) and (x_j^r, y_j^r, h). Then we can define the "spherical" visual ROI with center $((\theta_l+\theta_r)/2, (\phi_l+\phi_r)/2)$ and size $(|\theta_l - \theta_r|, |\theta_l - \theta_r|)$. This ROI corresponds to a rectangular ROI in unwarped cylindrical projection of the raw omnidirectional image. Thus, the image is unwarped on a cylinder where efficient image processing can be performed.

Each ROI can be used to compute visual descriptors of the detected object. These can be based, for example, on color distribution, intensity frequency analysis, or local motion analysis. If we denote the tuple of descriptors of blob j by $\mathbf{D}_j$, then a feature likelihood ratio can be computed as:

$$lr(D_j) = \frac{P_T(D_j)}{P_C(D_j)} = \frac{\gamma_T(D_j, \tilde{D})}{\gamma_C(D_j, \tilde{D})}$$

P_T and P_C are the probabilities that the descriptors have originated from target and clutter respectively. When a target track is initiated, a target descriptor $\tilde{D}$ is computed. It can either be held constant or updated during tracking (e.g. if light conditions change). Then P_T and P_C for descriptors D_j are computed using distributions γ_T and γ_C of D_j given the known a priori $\tilde{D}$.

We have experimented with a simple descriptor based only on color distribution $D(\mathcal{I}_j) = \{H(\mathcal{I}_j)\}$, where H_j is the color hue-saturation histogram of the image ROI $\mathcal{I}_j$ corresponding to blob j (the histogram is normalized and its mean value is subtracted from each component). If the correlation between two histograms is denoted by:

$$corr(H_i, H_j) = \frac{\sum_k H_i(k) H_j(k)}{\sum_k H_i(k)^2 \sum_k H_j(k)^2}$$

then the likelihood ratio is computed as follows:

$$lr(H_j) = \frac{\gamma_T(H_j, \bar{H})}{\gamma_C(H_j, \bar{H})} = \frac{max(corr(H_j, \bar{H}), 0)}{\frac{1}{1-corr_{min}}},$$

where $corr_{min}$ is a experimentally determined constant denoting the minimum value that the metric $corr$ can take (this is always a small negative number). From the equation above it is clear that negative values of $corr$ are treated as zero probability of detected target, while the probability of clutter detection is uniformly distributed (since the highest correlation between two histograms is one). The PDAF association probabilities are then augmented:

$$p_j = P_D e^{-\frac{1}{2}\tilde{\mathbf{y}}_j' S^{-1} \tilde{\mathbf{y}}_j} lr(H_j),$$

where $\tilde{\mathbf{y}}_j$ and S is the kinematic measurement residual vector and its covariance matrix, and P_D is a fixed detection probability [10]. It should be noted that color distribution is a very simple descriptor and is by no means sufficient for precise detection. While more sophisticated descriptors and detection probability distributions can be used, using color alone is inexpensive while still providing some level of target discrimination. We plan to extend the types of descriptors in future work.

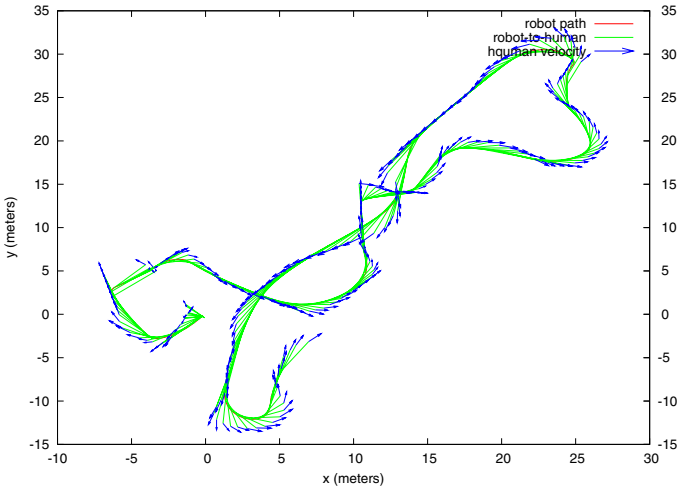

Fig. 4. A segment of a typical run: the solid curve represents robot positions, connecting to the corresponding estimated position of the human, the vector arrows show the person's estimated velocity (the plot shows only every 5-th estimate)

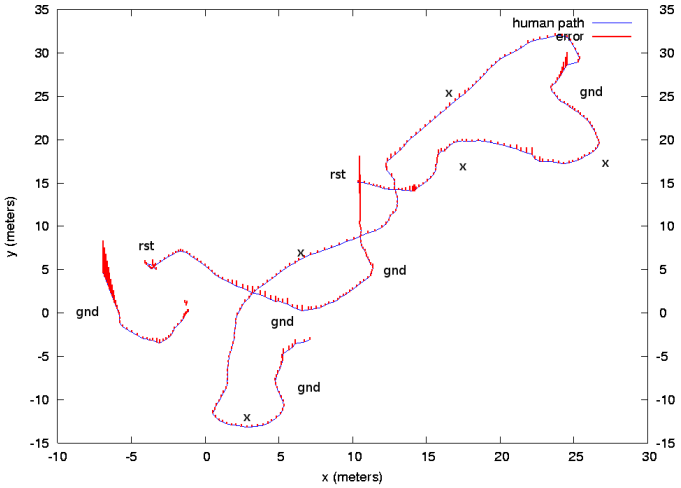

Fig. 5. Covariance norm of the target state estimate (every 3rd estimate drawn). A few typical situations during tracking: x (temporary target occlusions), gnd (the robot accelerates, pitches too much and sees mostly ground), rst (covariance test reset). In this case the tracker is able to successfully handle such conditions.

VI. Experiments

We conducted a number of experiments in a grassy park area with scattered trees and small bushes. We report results only for the second method presented since the first method often failed as described earlier. Fig. 4 shows a segment of a typical run during which the robot covers an area of 50x50 meters developing maximum speeds of 2.2 m/s. The person's estimated

velocity vectors are plotted as arrows. This particular run was chosen because it demonstrates several interesting situations that could happen during tracking. Fig. 5 shows the path of the human and the state estimate covariance norm as vertical bars for the same run. Since the RMP is dynamically balancing it pitches forward/backward when accelerating/decelerating. Often, during acceleration the laser on the robot can "loose" the human and see only the ground between the robot and the person. Such situation is marked as gnd in fig. 5. Whenever this happens, there are no valid measurements of the human, but there could be possibly measurements arising from clutter (i.e. nearby trees or bushes). Thus, the error of the estimate grows quickly and the tracker can either lock on a wrong target or it can simply continue tracking an invisible target with very large error. In our implementation, we have a maximum allowed error (usually 5 standard deviations of the measurement model) which will cause the robot to stop, and reinitiate the track. This is denoted by condition rst which occurs twice in fig. 5. Apart from ground effects, we had other people cross the path between the laser and human (marked with x), as well as walk together with the tracked person and then cross the path. In almost all cases the system was able to cope with such temporary occlusions (e.g. fig. 6). Of course, in a situation where an external person completely occludes the tracked one, for example, by walking right behind him for a long time, the system would quickly lock on the wrong one. In such cases, a more sophisticated set of visual descriptors must be used so that the tracker can reacquire the correct human after the occlusion.

We have used two different following controllers based on direct following and path following. The direct following controller minimizes the absolute relative bearing to the human and the absolute difference between the relative range and a desired following range. The path following controller is a pure pursuit type of path tracker which tracks the local path of the human at a desired following distance. Both controllers as well as the desired following distance account for the dynamic and kinematic constraints of the robot.

We measure the combined tracking and following performance based on the average following error, robot speed, and the range and bearing to target throughout a single run. The following error is computed as:

$$E_{follow} = r - d_{follow} - \frac{v^2}{2a_{min}},$$

where r is the range to human, d_{follow} is desired following distance, v is the current robot speed and a_{min} is the maximum robot deceleration. A small segment of a typical run is shown on fig. 7. For this case the following error is always within 0.5 meters and the bearing does not exceed 30^o. Average values for the four measures for 12 3-minute long runs are given in table I. The average following error in this set of runs was 30 cm at an average robot velocity of 1.5 m/s. During all runs there were a total of 21 rst conditions, 17 out of which the robot could handle by resetting on the correct target. 19 rst's were caused by the laser seeing the ground for at least 2-3 seconds and 2 by long occlusions caused by other people

Fig. 6. Output from the omnicam during tracking among trees: the robot is following the person in the middle (top frame - small circle indicates target position estimate), while another person (from the left) is crossing the path between the robot and the first person

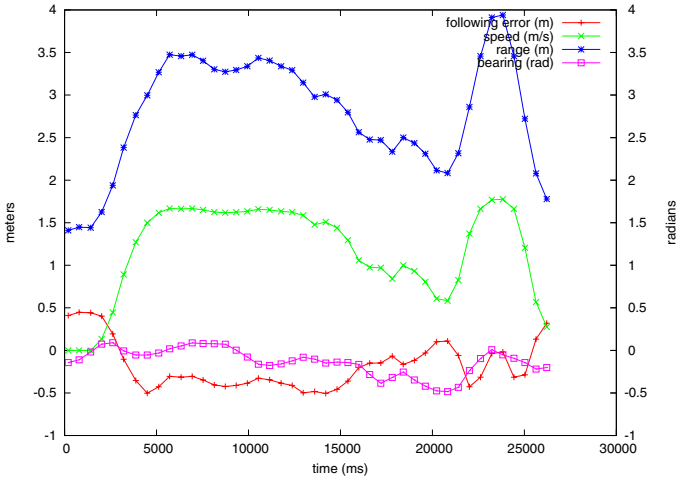

Fig. 7. A 30 sec. segment of a typical run at moderate speed. The following error is within 0.5 meters and the bearing to human is less than 30^o.

standing for too long in the path between the robot and tracked person.

VII. Conclusion

The paper presents two different methods for people tracking and following.

The first method suffers from failures because the egomotion compensation method used does not account for the actual structure of the environment and thus static nearby objects affect the frame differencing step. This is especially true at higher speeds among obstacles when clearly static objects violate the affine projection model assumption of the method.

Since the second method uses relative 3-D target position as measurements it is more robust in difficult environments as long as the target occlusions are short. In the future we plan to extend this system to use full depth measurement sensors such as a fast moving pan-tilt stereo head with wide field of view fish-eye lenses. Thus the laser blob extraction component can be replaced with a more robust 3D blob extraction based on depth and geometric human appearance templates (e.g.

r_{avg}	b_{avg}	v_{avg}	err_{avg}	v_{max}
2.91	0.22	1.30	0.70	2.17
4.04	0.16	1.49	0.30	2.25
3.99	0.14	1.49	0.23	2.16
3.74	0.18	1.44	0.15	1.67
3.56	0.15	1.42	0.15	1.77
5.11	0.14	1.72	0.40	2.30
5.46	0.12	1.78	0.44	2.27
3.29	0.14	1.32	0.18	1.71
4.39	0.15	1.56	0.30	2.41
3.91	0.14	1.49	0.24	2.06
4.40	0.12	1.59	0.24	2.06
3.66	0.16	1.38	0.27	1.91
4.04	0.15	1.50	0.30	2.06
0.72	0.03	0.15	0.15	0.25

TABLE I

AVERAGED ESTIMATES FOR 12 RUNS OF 2-3 MINUTES EACH: r_{avg} - RANGE; b_{avg} - ABSOLUTE BEARING; v_{avg} - ROBOT VELOCITY; err_{avg} - ABSOLUTE FOLLOWING ERROR; v_{max} - MAXIMUM ROBOT VELOCITY. THE BOTTOM TWO ROW SHOWS THE MEAN AND STANDARD DEVIATION OF THE FOUR MEASURES OVER ALL 12 RUNS.

[17]). 3D kinematic data combined with more sophisticated conjunctions of visual cues and constraints (e.g. [6]) would constitute a powerful set of features for data association. A tracker based on such a framework would be more robust to clutter, occlusions, and noisy environments. We plan to adopt these methods and improve the performance of the presented system in our future work.

REFERENCES

[1] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Minerva: A second generation mobile tour-guide robot," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'99)*, 1999.

[2] J. F. Engelberger, "Health-care robotics goes commercial: The helpmate experience," in *Robotica*, vol. 11, 1993, pp. 517–523.

[3] L. Davis, V. Philomin, and R. Duraiswami, "Tracking humans from a moving platform," in *ICPR '00: Proceedings of the International Conference on Pattern Recognition (ICPR'00)-Volume 4*. Washington, DC, USA: IEEE Computer Society, 2000, p. 4171.

[4] B. Jung and G. S. Sukhatme, "Detecting moving objects using a single camera on a mobile robot in an outdoor environment," in *International Conference on Intelligent Autonomous Systems*, Amsterdam, The Netherlands, Mar 2004, pp. 980–987.

[5] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," *Comput. Vis. Image Underst.*, vol. 73, no. 3, pp. 428–440, 1999.

[6] C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 560–576, 2001.

[7] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *Int. J. Comput. Vision*, vol. 39, no. 1, pp. 57–71, 2000.

[8] I. J. Cox and S. L. Hingorani, "An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, 1996.

[9] Y. Bar-Shalom and X. Li, *Multititarget-Multisensor Tracking: Principles and Techniques*. Stoors, CT: YBS, 1995.

[10] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[11] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "People tracking with mobile robots using sample-based joint probabilistic data association filters," *IJRR*, vol. 22, no. 2, 2003.

[12] T. B. J., S. W. Hart, D. R. Karuppiah, J. D. Sweeney, , and O. Brock, "Cascaded filter approach to multi-objective control," in *IEEE International Conference on Robotics and Automation*, New Orleans, USA, April 2004, pp. 3877–3882.

[13] C. Geyer and K. Daniilidis, "Paracatadioptric camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 687–695, 2002.

[14] B. Micusík and T. Pajdla, "Omnidirectional camera model and epipolar geometry estimation by ransac with bucketing." in *SCIA*, 2003, pp. 83–90.

[15] S. Baker and S. K. Nayar, "Single viewpoint catadioptric cameras," in *Panoramic Vision: Sensors, Theory, Applications*, R. Benosman and S. B. Kang, Eds. Springer-Verlag, 2001.

[16] J. Gluckman and S. K. Nayar, "Ego-motion and omnidirectional cameras," in *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998, p. 999.

[17] D. Beymer and K. Konolige, "Real-time tracking of multiple people using continuous detection," in *International Conference of Computer Vision*, 1999.