

A Laser-Based People Tracker

Ajo Fod

Andrew Howard

Maja J Matarić

Computer Science Dept.

University of Southern California

Los Angeles, CA 90089-0781, USA

fod@fnord.usc.edu, {ahoward|mataric}@cs.usc.edu

Abstract—

This paper describes a method for real-time tracking of people in everyday environments, using multiple planar laser range-finders. People tracking is a well-studied problem in machine vision; in this paper, we adapt some of those methods to laser range-finders. We group range measurements into entities such as blobs and objects, and use a Kalman Filter to estimate trajectories for these objects. The filter is able to generate smooth trajectories, even when objects are occluded. This paper presents our evaluation of the tracker's performance in a series of four experiments.

Keywords—People tracking, laser sensor, Kalman filter, occlusions.

I. INTRODUCTION

Tracking people, for surveillance, crowd control; activity recognition and characterization, etc., is a well studied problem in machine vision (e.g., [19]). In this paper, we present a method for tracking people using an alternative sensor: planar laser range-finders. Laser-range finders have rapidly gained popularity in the robotics community, where they have been used for applications such as collision avoidance, localization [9], [26], and map building [25], [11]. In this paper, we demonstrate that lasers can also be used to effectively track people.

As sensors, lasers have both advantages and disadvantages when compared with vision. They are able to reliably generate range measurement that are accurate to within a few millimeters over distances of many meters, but have limited angular resolution (around 0.25°). They are easy to calibrate and insensitive to lighting conditions and other external disturbances, but cannot provide information such as the color of objects in the environment. Lasers have a very wide field-of-view (which means that relatively few are required to cover a workspace), but they are much more expensive than cameras.

Earlier work on motion tracking for crowded environments using scanning laser range finders was done by Prassler *et al* [18], who used occupancy grids and linear extrapolation of occupancy maps to estimate trajectories. In contrast, our methods involve the use of multiple lasers and trajectory estimation algorithms. We study the problem of tracking multiple moving objects in a workspace covered by multiple lasers. We draw from vision-based approaches to estimate motion recursively from subsequent scans of the laser. We use a version of the constraint formulation for

rigid bodies [22], [23] and an abstract representation of objects for tracking. We associate a Kalman Filter (KF) with each object to alleviate the consequences of occlusions and to reduce the impact of occlusions and model inaccuracies.

In this paper, we present four evaluation experiments to measure the performance of the tracker. Three experiments were performed to measure the accuracy with which the tracker estimates the position of objects in the workspace, and the fourth to measure the accuracy in estimating the number of objects in the workspace. The first experiment compares the scatter in position estimates of large objects in comparison with a beacon of known diameter. The second shows that the tracker is stable and that no errors develop over time. The third experiment measures the residual of the KF while estimating the position of objects. The fourth experiment shows that the tracker accurately estimates the number of object being tracked.

II. APPROACH

A. Overview

The laser returns range measurements to the nearest obstacle, in polar coordinates. From this we construct a model of the process that generates the range readings. The model consists of two parts: the *Background Model* and the *Foreground Model*. The Background Model is used to filter background information that is not of immediate relevance to tracking. The Foreground Model is used to predict range measurements that are not explained by the Background Model.

B. Background Model

Background models used in most vision applications utilize pixel-level statistics [15], [19]. Each new pixel value is classified as the background or a moving object. We adapt this technique for use with laser scanners. Since a laser range measurement gives us the range to an object, we can use a modified rule for the classification of foreground and background. We assume that the farthest known stationary object is part of the background. In order to be robust to small errors, we also maintain the mean and the variance of the measurements that are classified as background and use the information to test the hypothesis that any reading is part of the background.

C. Foreground Model

The tracking process is divided into two subtasks: 1) tracking blobs and 2) grouping surfaces into objects. A *BlobTracker* tracks the blobs, and an *ObjectTracker* estimates paths of collections of blobs, called objects, in the workspace.

C.1 Blobs

We define a *Blob* as a grouping of adjacent foreground readings from the laser that appear to be on a continuous surface. The set of measurements that form this continuous surface constitute the *signature* of the blob. Since the laser is oblivious to the component of motion perpendicular to the laser's plane of scan, we do not attempt to model motion or changes along this direction, i.e., we only consider planar motion. Human beings are in most cases continuously observable in a plane parallel to the ground at about waist-height in planar workspaces. Thus, we chose this plane for laser tracking. In our experiments, we assume that measurements that are spatially separated by less than 10cm belong to the same blob.

The Foreground model contains two generations of blobs at any given instant: **gen-x** - the old blobs and **gen-y** - the new blobs. To reduce the search space in matching blobs for state reestimation, we draw a rectangular bounding box around any blob. Assuming a bound on average acceleration of $10m/s^2$ (about 1g) and scan frequency of 5 Hz, we estimate the maximum error in the prediction of the state to be about 50mm. A match is said to exist between points in the prediction of the position of a gen-x blob and gen-y blob if the distance between them is less than 50mm. We attempt to pair every point in gen-y with two points in gen-x that match, and thus have the minimum distance to the new point. We use these two points for a linear approximation of the surface of the blobs using the chamfer method suggested in [1].

Newly discovered gen-y blobs inherit their state vector as a linear combination of the predicted state vectors from the matched gen-x blobs. The weighting of each of the gen-x blobs in determining the state of the gen-y blobs is given by the number of point matches between the blobs. Blobs that do not have any parents are initiated with a state vector of 0. The gen-x blobs are removed and the current gen-y blobs become the gen-x blobs for the next scan.

Many methods exist for determining the structure of objects independent of motion [16], structure from motion [13], [17], and motion of a known structure [24], [6], [7]. Koller [14] described a method of tracking objects using a template of the object. We use a rigid body model with the signature from the previous scan and a velocity state estimated from earlier scans to predict the measurements from the next scan. The matched blob's signature is used to update its rotational and rectilinear velocities.

We modify the method of subspace filtering as presented

in [22], [23] for blob state estimation. Soatto's method requires one-to-one correspondences between features. It is not easy to find such correspondences in the measurements from the laser. In laser sensor space, our strength is in the ability to determine the range reading accurately - so we try to fit depth maps. We approximate the surface of objects to a sequence of lines connecting predictions of the signature of a blob. Each new measurement is paired with the two closest adjacent predictions in the new scan. We use the distance between the line joining the predictions and the closest point as a measure of the prediction error. This error is summed over all points in the new surface to obtain a metric of prediction performance.

C.2 Objects

In our approach, objects map to multiple blobs. For example, the hands and the torso may form different blobs. These are unified into one percept of an object by the *ObjectTracker*. Intille et al., [21] describe heuristic methods to explain the scenes described by blob trackers for vision. Although our algorithm is based on some of these principles, because of the accuracy of range measurements we do not need to distinguish between blobs that occupy the same space, since we can identify them easily.

Each blob inherits its *ObjectTracker* from its parent in the previous generation. When two blobs form the parent of a blob, the closest *ObjectTracker* is associated with the new blob. Blobs that appear spontaneously within a prescribed radius of any *ObjectTracker* are associated with it. Whenever no known *ObjectTracker* is sufficiently close to a new blob, a new *ObjectTracker* is constructed for the blob. Once instantiated, *ObjectTrackers* are allowed to persist until supporting blobs are not visible for over 5 seconds - an empirically determined threshold.

When objects move out of the plane of the laser, the *ObjectTracker* is left with no blobs to support it. It continues scanning for objects that it can describe with fewest errors. For the duration of the occlusion, state updates are performed as suggested in [8], [19]. When the object reappears, the *ObjectTracker* becomes active again, as if the object had never disappeared.

The tracker we implemented was designed to work in real time. The rigid body component of the motion can be determined after approximately five consecutive laser scans. Whenever the delay in processing exceeds a certain threshold, we skip frames to speed up the processing. With these modifications, we can track up to 20 objects simultaneously, using 6 laser scanners, without significant degradation of performance.

III. EXPERIMENTAL PROCEDURE

Our experiments were based on real laser scan sequences. In this Section we report our observations from experiments conducted on data obtained in our lab. We

used Player [10], a server and protocol that connects robots, sensors and control programs across the network to communicate with our laser range-finders.¹ We used standard SICK planar scanning laser range-finders placed 0.75 – 0.80m above ground level, targeting the waist of an average-height walking person. This is the cross-section which moves least. Moreover, people rarely move out of this plane. For example, a person who sits is still visible, albeit at shoulder height.

We performed four sets of experiments to measure the performance of the tracker. Three experiments were performed to measure the accuracy with which the tracker estimates the position of objects in the workspace, and the fourth experiment to measure the accuracy in estimating the number of objects in the workspace.

A. Apparatus and Methods

We made two sets of measurements in each experiment set. The first measurement - the ground truth - was recorded using a camera and post-analyzed by a human operator. The second was the output from the tracker. The experiments were performed with multiple lasers in a single room. A single room was used in order to make the measurement of ground truth easier.

B. Experiment 1: Positioning Accuracy

We performed the following two tests to determine the accuracy of the tracker's estimate of object position. The first set used stationary beacons whose position in the laser scan can be accurately determined. This was done to measure the performance of the tracker in the absence of any noise arising out of factors like the shape, size, and movement of human beings. The second set was performed with human subjects. The goal was to measure the performance of the tracker in the presence of such noise.

We obtained two measurements of position for each sample point: one in the laser frame of reference, and the other in the visual frame of reference (henceforth frame of reference is abbreviated to frame). Figure 1 shows the room used in the experiment. In order to get measurements in the visual frame, we positioned beacons at the corners of the 2' x 2' floor tiles. A second set of measurements of the positions of the beacon was made in the frame of the laser.

We determined the transformation T from the laser frame to the visual frame which minimized the mean squared distance between corresponding points in the two frames. This transformation is used subsequently to map the position of any object in the laser frame to the corresponding point in the visual frame. The beacon positions are shown in Figure 2. The error or positional inaccuracy

¹Player was developed jointly at the USC Robotics Research Labs and HRL Labs and is freely available under the GNU General Public License from <http://playerstage.sourceforge.net>.

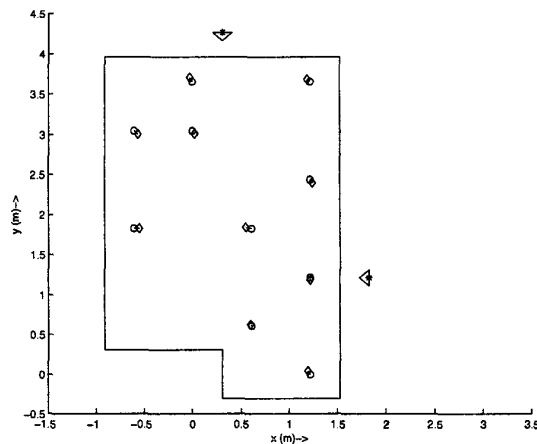


Fig. 1. **Experiment 1: DUMMY FIG.** An image of the room used in the experiments.

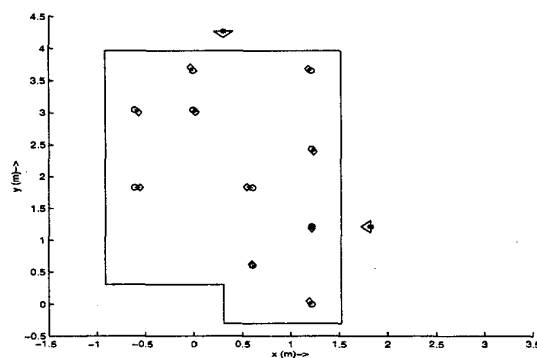


Fig. 2. **Experiment 1:** The positions of the beacons in the room. The diamonds show the location of the beacon in the laser frame and the circles in the visual frame. Correspondences are connected with a line. The two lasers are marked with a "*" and the arrow shows their direction of scan.

can be plotted as a set of 2D displacements from the position measured in the visual frame to the position measured in the transformed laser frame. In the ideal case of zero error, these vectors would all be located at the origin. The error plot is shown in Figure 3.

We performed a second set of measurements of people standing on tile corners. For this set, the subjects moved between tile corners and waited for about a second at each position. We used the transformation T determined using the beacons to determine the error in the position estimate measured by the laser for the human movement data. In Figure 4, the scatter plot for this set of measurements shows that the error is approximately the size of the cross-section of a human torso: about 0.3m.

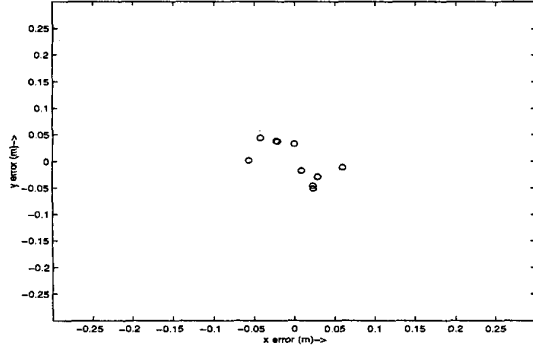


Fig. 3. **Experiment 1:** Each point represents the displacement required to map the measurement of the beacon in the laser frame to the visual frame. This shows the error vector for position estimates made by the tracker.

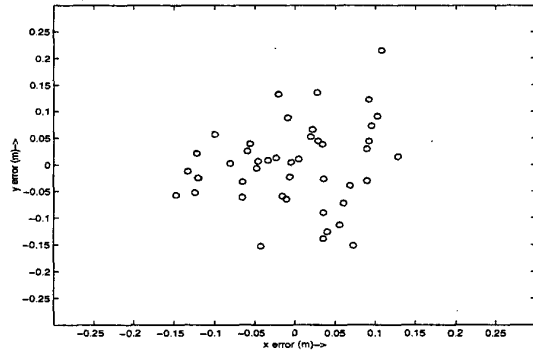


Fig. 4. **Experiment 1:** The plot shown in Figure 3 is repeated with people in this figure to show the change in the scatter distribution. Notice that it is now approximately the size of the cross-section of a human torso.

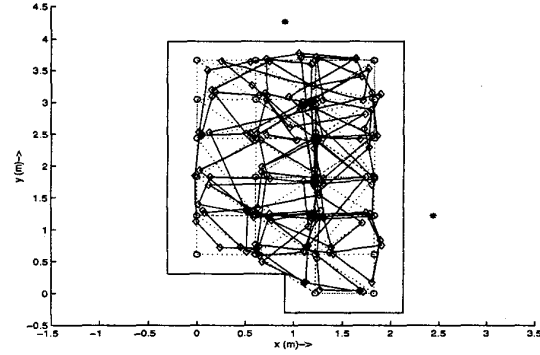


Fig. 5. **Experiment 2:** Paths in the laser frame (dotted lines) superimposed over paths in the visual frame (solid lines).

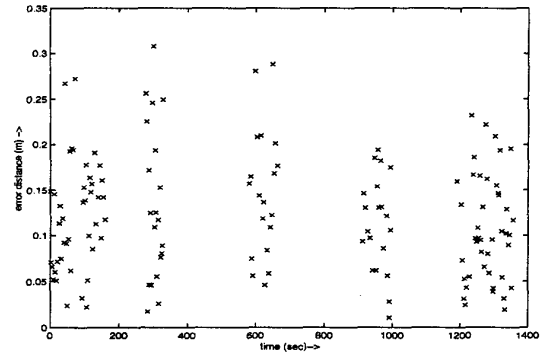


Fig. 6. **Experiment 2:** The magnitude of the error is shown here as a function of time for individual measurements. These measurements were made at 40 points separated by 5 minutes. No points were plotted where measurements were not made. The figure shows that the drift does not increase significantly over a 20 minute period.

C. Experiment 2: Drift in Tracking

We performed the following experiment to determine if any cumulative error in the position measurements developed over time. We obtained a set of measurements in the same workspace, along the lines of the second test in Experiment 1, for a duration of 20 minutes. At least 40 correspondences between the laser and the visual frames were recorded starting at every 5 minute interval. Figure 5 shows a superposition of the paths followed by a person in the visual frame with those in the laser frame. Figure 6 shows a plot of the positional error as a function of time. The figures demonstrate that there is no noticeable drift or systematic increase in RMS error over the 20 minute period.

D. Experiment 3: Dynamic Performance

To measure the performance of the KF in estimating the state of the object, we developed an error metric which measures the error in predicting the trajectory of the object. We defined two quantities for every object - $\Delta \bar{\mathbf{c}}_t$: the observed change in position and $\Delta \hat{\mathbf{c}}_t$: the predicted change

in position at time t . We define the error metric to be:

$$\epsilon_1 = \frac{1}{T} \sum_{t=1}^T |\Delta \bar{\mathbf{c}}_t - \Delta \hat{\mathbf{c}}_t|^2 \quad (1)$$

where ϵ_1 is the residual, a measure of the instantaneous uncertainty or variance in the error of the position estimate of the blob. The variance was in the range $1 - 5 \text{ cm}^2/\text{s}$. As expected, it was larger for objects that moved around (e.g., people) and smaller for objects that were stationary.

E. Experiment 4: Tracked number of objects

We performed the following experiment to determine the accuracy of the tracker in determining the number of people in the room. Tracking was performed on a scripted play (shown in Figure 7) which involved normally observed activities in the room such as walking, talking, group discussions, and path crossings.

The tracker can only see or track objects that have moved and cannot differentiate between people and objects. Thus

time	action
0.00	Start of program - with A at <i>Station A</i>
1:33	A moves fully (initialization)
1:40	B walks into the room, goes to <i>Station B</i> , and sits down
1:56	A gets up and walks over to <i>Station B</i> to talk to B
2:15	C walks in for a cup of coffee to the <i>coffee table</i>
2:25	C walks back out with cup
2:40	D walks in and shakes hands with A
2:49	D pulls up a chair next to <i>Station A</i> and sits down
2:58	A goes to <i>Station A</i> and sits down to talk
...	...
7:03	A walks to one of the moved chairs
7:10	B walks in towards the <i>printer</i>
7:12	A and B cross paths
7:15	A bends below the laser's range
7:19	A gets back in the range
7:27	A walks back to <i>Station A</i>

Fig. 7. Experiment 4: Excerpts from the script we used.

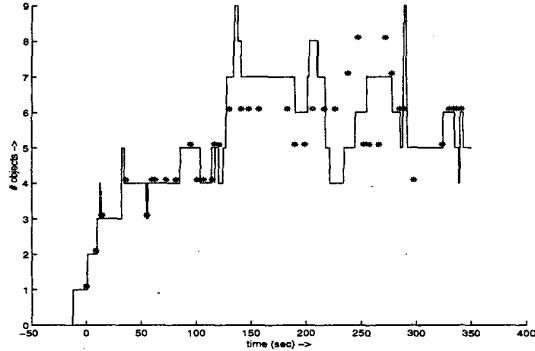


Fig. 8. Experiment 4: Plot of accuracy in tracking the number of moving objects in the room. Stars represent the number of objects as measured by the camera and solid line the number tracked by the tracker.

we used the visual ground truth measurement to find the number of objects that moved. Figure 8 shows the number of objects that moved as measured by the tracker compared with the visual measurements.

E.1 Results

The errors in estimating the number of objects in the room were analyzed with the aid of the script and the video. ObjectTrackers are designed to track objects that have moved once. Consequently, chairs appear spontaneously near previously existing objects (people). While the tracker retains separate identities for people and chairs, when people are sitting on chairs, it becomes difficult at times (as when the chair is temporarily occluded), to continue to distinguish between the chair and the person. In these cases, the tracker underestimates the number of objects in the room. This phenomenon occurred between 220s and 270s in the Figure 8. On the other hand, the number of objects in the room is overestimated when they move very quickly or when they jump, and the KF is unable to keep up. This sometimes results in temporary objects and

is manifested in spikes at about 150s, 220s, and 290s in Figure 8. Occasionally, the newly spurious object becomes the parent of a stable blob. When this happens, the number of objects is overestimated for an extended period of time, as between 170s and 200s, in the same figure.

IV. DISCUSSION

In this paper we introduced a laser-based tracking approach. Our method exploits the advantages of lasers while overcoming some of the shortcomings of vision. Unlike vision, where multiple cameras are required to cover a scene, a single laser can effectively cover a reasonably large workspace. Lasers provide a "gist" of the scene in a plane very accurately, with low computational overhead.

Our tracking algorithm is computationally light; it works in real time on an IBM 600e Think-pad with 64MB RAM and a 300MHz P-II processor, with 2 lasers and up to 10 objects. We have also collected extensive data with a 900MHz P-III with 128MB RAM, using up to 6 lasers covering two rooms.

Using lasers has its drawbacks. It is difficult to obtain a set of features that uniquely distinguish one object from another. While it is possible to use the velocity and position of objects at the start of an occlusion in order to reason about their location after the occlusion is removed, it is not easy to do so conclusively. Lasers are oblivious to objects outside their plane of scan, thus making it difficult to perceive details of activity. It is possible to use multiple lasers in different planes to build up a 3D representation of the scene. Multiple lasers can also be used in large workspaces with a many moving people, where occlusions prohibit effective tracking. As a practical matter, lasers are currently expensive relative to cameras.

A. Applications

Gait recognition and activity modeling are among the interesting applications of trackers that can store and process entity-specific information. Bregler and Brand [4], [2], [3] use motion capture to record joint orientation data for movement analysis. One could modify these to use laser data to determine patterns in the movement in a plane, possibly to identify the task being performed or the person performing it.

Microsoft's research on EasyLiving Technologies [12], [5] is a commercialization of tracking technology which requires locating and tracking individuals in a closed workspace. The goal is to guess the intent of the user and help in automating some of the mundane tasks like switching appliances on in the presence of a user, setting parameters like the temperature in a workspace automatically, or even logging on automatically in the presence of a user.

Robotic security devices [20] need to identify moving objects and possibly determine helpful information about them. Useful features include the number and type of ob-

jects and their physical characteristics. These features can potentially be determined with a laser tracker.

In our tracking work, we filtered out the background. However, we could instead filter out the foreground. Moving the laser would then provide us with an estimate of the velocity of the sensor. We have used the KF to estimate a differential state but could instead use it to estimate the displacement directly. This would be applicable to laser-based map building, navigation, and localization, as suggested in [11], [9], [26].

V. CONCLUSION

We have shown that lasers are an effective sensor for object and people tracking by developing a computationally efficient algorithm for multiple object tracking. We validated the tracker in a series of four experiments in a laboratory workspace, and demonstrated its effectiveness in the presence of occlusions.

In our continuing work, we are scaling up the presented approach to track large groups of interacting humans in complex environments with multiple lasers.

ACKNOWLEDGEMENTS

The research described here was supported in part by the NSF Career Grant IRI-9624237 to M. Matorić, in part by ONR DURIP Grant N0014-00-1-0140 for equipment support, and in part by DARPA MARS Program Grant DABT63-99-1-0015.

REFERENCES

- [1] H. G. Barrow. Parametric correspondence and chamfer matching: Two new techniques for image matching. In *Proc. Fifth International Joint Conference on Artificial Intelligence*, volume 2, pages 650–663, Cambridge, MA, 1977.
- [2] M. Brand. Understanding manipulation in video. In *Proc. 2nd International Conference on Face and Gesture Recognition*, pages 94–99, Killington, VT, 1996.
- [3] M. Brand. Style machines. In Kurt Akeley, editor, *Proc. Computer Graphics Siggraph 2000*, pages 183–192, New Orleans, Louisiana, USA, July 23–28 2000. ACM Press / ACM SIGGRAPH / Addison Wesley Longman.
- [4] C. Bregler and J. Malik. Learning appearance based models: Mixtures of second moment experts. *Advances in Neural Information Processing Systems*, 9:845, 1997.
- [5] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer. Easyliving: Technologies for intelligent environments. In *Proc. Handheld and Ubiquitous Computing*, pages 12–29, Bristol, September 2000. Springer Verlag.
- [6] D. Gennery. Tracking known 3-dimensional objects. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13–17, Pittsburg, PA, 1982.
- [7] D. Gennery. Visual tracking of known 3-dimensional object. *Int. J. of Computer Vision*, 7(3):243–270, 1992.
- [8] D. Koller, J. Weber, and J. Malik. Robust multiple car tracking with occlusion reasoning. In *Proc. Third European Conference on Computer Vision*, LNCS 800, pages 189–196. Springer-Verlag, 1994.
- [9] D. Fox, W. Burgard, and S. Thrun. Markov localization for mobile robots in dynamic environments. *Artificial Intelligence*, 11:391–427, 1999.
- [10] B. Gerkey, R. Vaughan, K. Sty, A. Howard, G. Sukhatme, and M. Matorić. Most valuable player: A robot device server for distributed control. In *Proc. IEEE International Conference on Intelligent Robots and Systems (IROS 2001)*, pages 1226–1231, Wailea, Hawaii, 2001.
- [11] J. Gonzalez, A. Ollero, and A. Reina. Map building for a mobile robot equipped with a laser range scanner. In *Proc. IEEE Int Conference on Robotics and Automation*, volume 3, pages 1904–1909, San Deigo, CA, USA, May 1994.
- [12] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *Proc. IEEE International Workshop on Visual Surveillance*, pages 3–10, Dublin, Ireland, July 2000.
- [13] J. Oliensis, J. Inigo, and Thomas. Recursive multi-frame structure from motion incorporating motion error. In *Defense Advanced Research Projects Agency (DARPA) Image Understanding Workshop*, pages 507–513, San Diego, CA, 1992.
- [14] D. Koller. Moving object recognition and classification based on recursive shape parameter estimation. In *Proc. 12th Israel Conference on Artificial Intelligence, Computer Vision*, pages 359–368, Ramat Gan, Israel, December 27–28 1993.
- [15] K. Toyama, J. Krumm, and B. Brumitt. Wallflower principles and practice of background maintenance. In *Proc. International Conference on Computer Vision*, pages 255–261, Corfu, Greece, Sep 1999.
- [16] M. Lee, G. Medioni, and R. Deriche. Structure and motion from a sparse set of views. In *Proc. IEEE International Symposium on Computer Vision*, pages 73–78, Coral Gables, Florida, USA, November 1995.
- [17] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter based algorithms for estimating depth from image sequences. *Int. J. of Computer Vision*, 3(3):209–238, 1989.
- [18] E. Prassler, J. Scholz, and A. Elfes. Tracking people in a railway station during rush hour. In H.I. Christensen, editor, *Proc. Computer Vision Systems, First International Conference, ICVS'99*, pages 162–179, Las Palmas, Spain, 13–15 Jan 1999. Springer-Verlag.
- [19] R. Rosales and S. Sclaroff. Improved tracking of multiple humans with trajectory prediction and occlusion modeling. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Workshop on the interpretation of Visual Motion, Santa Barbara, CA, 1998.
- [20] Commander H. R. Everett (Rtd). Breaking down the barriers. *Unmanned Vehicles*, 3(1):18–20, Feb-Apr 1998. also at: www.nosc.mil/robots/land/robart/history.html.
- [21] S. Intille, J. Davis, and A. Bobick. Real-time closed world tracking. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 697–703, Puerto Rico, June 1997.
- [22] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. In *Proc. of the 33rd IEEE Conf. on Decision and Control*, pages 61–72, Lake Buena Vista, Florida, Dec 1994.
- [23] S. Soatto and P. Perona. 3d visual motion estimation using subspace constraints. In *Proc. of the 1st IEEE Conf. on Image Processing*, 1994.
- [24] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE Trans. Pattern Analysis Machine Intelligence*, 8(1):90–99, Jan 1986.
- [25] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation. *Artificial Intelligence*, 1:21–71, 1999.
- [26] Sebastian Thrun, Dieter Fox, Wolfram Burgard, and Frank Dellaert. Robust monte carlo localization for mobile robots. *Artificial Intelligence*, 128(1-2):99–141, 2001.