Laser Motion Detection and Hypothesis Tracking from a Mobile Platform

Joshua Bobruk[1] **David Austin**[1,2]
josh@syseng.anu.edu.au d.austin@computer.org

[1]Robotic Systems Lab, RSISE
Australian National University,
ACT 0200, Australia

[2]National ICT Australia,
Locked Bag 8001,
Canberra, ACT 2601

Abstract

A method to detect and track moving objects with a laser range scanner from a mobile platform is presented. Detection and tracking is achieved by noticing differences in laser scan data over time, and maintaining hypotheses which reflect the chance of the changes being caused by moving objects. A general model of movement is employed to allow the tracking of multiple types of objects, from humans to other moving platforms, in one system. Pure odometry data and scan matching is used to compensate for motion, and hypotheses are maintained using a Kalman filter to estimate the velocity of objects. The approach is employed to track objects in a range of environments, where cameras mounted on an active head are used to record the any movement around the platform.

1 Introduction

The ability to track moving objects, from a moving platform is a skill required in a number of fields from surveillance [Siebel and Maybank, 2002] to robot human interaction [Feyrer and Zell,]. Methods exist for detecting targets with a range of sensors, however many are only designed to track one particular type of target [Beymer and Konolige, 1999], and many fall victim to a severe drop in accuracy as the target move towards the extremes of the sensors. Vision systems can provide a large amount of detail about a subject, but this can be overwhelming, and once a target moves too close, no further useful information can be obtained. Laser range scanners, although slightly more expensive than other sensors, provide a high accuracy map of the environment, and perform well when objects are close or far from the scanner with ranges of up to 80m. The data is limited to a plane and objects can become occluded, however, this is acceptable for the task of tracking moving objects where accuracy over a wide range is of a greater importance.

Rather than tailoring a solution to solve the specific problem of tracking humans, this method provides a more general approach that takes any moving object as an object of interest. As such, this method can be used to simultaneously track a number of different types of target in a given scene. By remaining general, the approach could easily be used for applications where other mobile platforms need to be detected and perhaps interacted with. For this reason, techniques such as signature matching [Fod *et al.*, 2002] for detecting objects, or vision systems matching templates [Needham and Boyle, 2000] have been avoided. The preferred method is to track objects with a minimal amount of data, with no filtering based on templates to avoid losing important information about the state of the scene.

Methods have been presented to track moving objects using systems such as particle filters [Schulz *et al.*, 2000] and similar techniques [Tao *et al.*, 1999]. These are shown to work well in their respective applications, however, for the task of tracking objects from a mobile platform, more control of tracking was needed. The task of tracking from a mobile platform becomes more difficult than the stationary case as the side effects of egomotion need to be compensated. By using an hypothesis tracking method, an keeping hypotheses of where objects might be moving in the scene, we can easily modify the weightings of the given hypotheses based on the robot motion and occlusions that may occur as platform becomes mobile.

2 Outline

The structure of this document presents the overview of the system used to solve this problem in Section 3, followed by the breakdown and description of each of the elements in Sections 4 to 9. At the end of the document, results are provided showing the effectiveness of the proposed method.

3 Approach

The goal here is to take laser scans at any given height from a mobile platform, and process the data until moving targets in the scene can be detected. The experiments outlined here were all conducted on an XR4000 Robotic platform shown in Figure 1.

The laser range scanner used is the Sick Laser Range Scanner, and an example of a raw laser scan obtained from this device can be seen in Figure 2. Once the scan has been taken, a number of steps are required to process the data to obtain areas where motion is actually occurring as opposed to simply artifacts of noise. If the scan data is not processed, any noise will result in the misdetection of objects resulting from the given noise. The overall design of the system proposed can be seen in Figure 3.

The design of the system tries to solve the major issues involved with tracking a range of objects simultaneously from a moving platform. These issues are:

- Noise from laser scanner and odometry
- Time stamping of data
- Occlusions as the platform moves around objects

We attempt to solve these problems by first finding changes in the scene that could represent motion. To find changes in the scans, the scans are first converted to an occupancy grid map format. The grid map has been used in multiple approaches to motion detection [Schulz *et al.*, 2000], and is a natural way to approach the problem. Grid maps convert the raw laser scan data (see Figure 2) into a square grid of occupied cells surrounding the scanner. A typical grid map can be seen in Figure 4.

It is clear to see that when objects move around the scanner, the moving objects will cause changes in the scan data over time, and subsequently the grid maps will indicate where moving objects occur. When the laser scanner is stationary, finding moving objects is a simple task of finding the differences in two subsequent maps (see Section 5), however, when the scanner is moved,

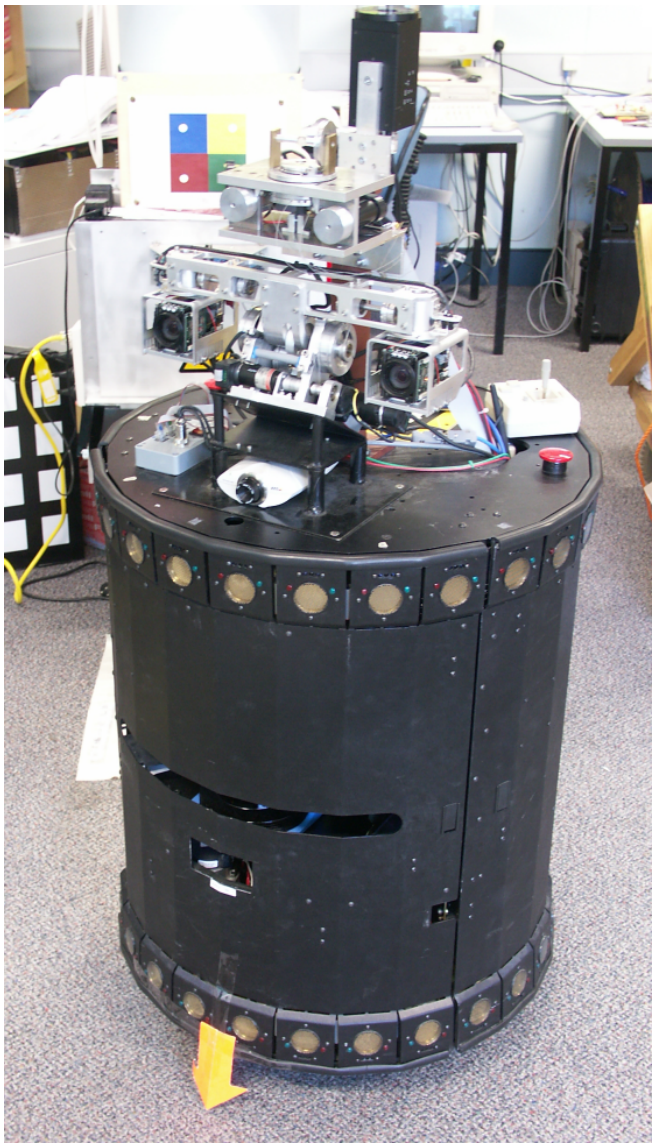

Figure 1: XR4000 robotic platform used in the experiments outlined.

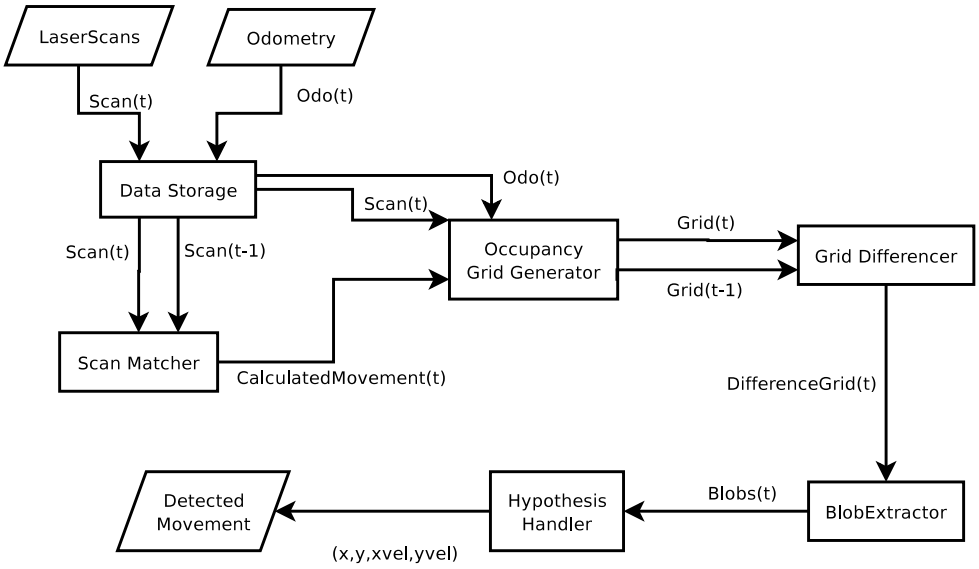

Figure 3: Top level design diagram showing the processes involved in tracking moving targets using an hypothesis method and laser scan data.

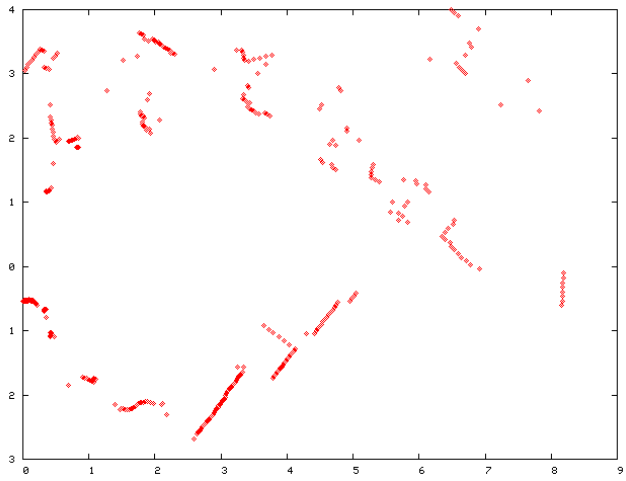

Figure 2: Example of Laser Data in Cartesian Coordinates with the scanner at (0,0)

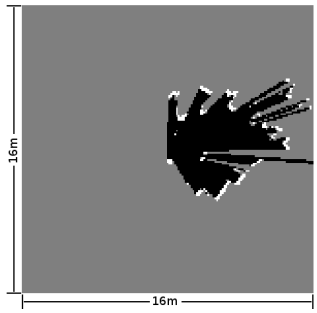

Figure 4: Laser Scan converted into Grid Map format with the scanner situated in the centre of the grid.

compensation is required to remove the effects of platform rotation and translation.

The compensation is obtained by combining odometry data and scan matching techniques, resulting in grid maps where structure remains almost stationary despite platform movement. The scans are then differenced leaving a grid containing all the changes between the two grids. A typical difference map and corresponding grid maps can be seen in Figure 5.

The difference map is then segmented into *blobs* of connected cells, and each blob is then analysed to determine whether it should be contributing to a constant velocity motion hypothesis. By choosing a constant velocity model, we can track a wide variety of targets independent of scanner placement, or object shape.

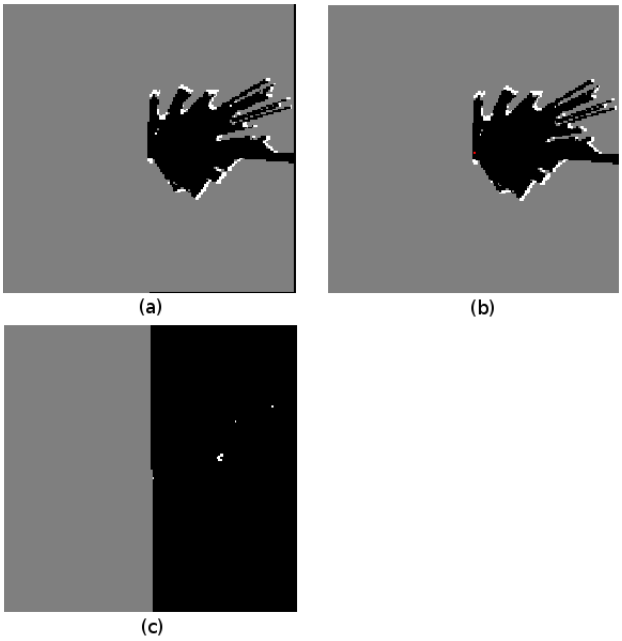

Figure 5: Two subsequent scans and the resulting difference map. (a) Current grid map (b) Previous grid map (c) Difference result with white regions representing changes, and grey regions representing space with no valid data.

Hypotheses at this stage are maintained using a Kalman filter for each hypothesis and contain useful information about the position and velocity of moving objects, and if a given hypothesis is maintained for a number of samples, it is deemed to be a moving object, and therefore of interest.

4 Motion Compensation

The removal of ego motion was initially attempted by using only odometry data from the mobile platform. This proved to be largely inconsistent with the observed laser data due to delays and timing errors. To compensate for this, each scan was matched against the previous scan using scan matching techniques [Lu and Milios, 1994] [Besl and McKay, 1992]. In this method, the scan matching method used is an iterative closest point algorithm which returns a calculated rotation and translation to match each scan with the previous scan. This works well in most cases, and only fails in scans with very little structure in the scanned environment. When combined with odometry data, and as long as the platform is limited to a fixed maximum rotation and translation speed, the scans will match well, giving close estimates to the amount moved by the platform.

5 Occupancy and Difference Maps

Once the motion of the platform is determined, the latest scan can be compensated and converted to an occupancy grid map. As we have seen, grid maps consist of an $n \times n$ square grid where cells contain information about the occupation of regions around the scanner which is positioned at the centre of the grid. The value of each cell represents the probability that it is occupied between 0 and 1 (See Figure 4).

For this approach, it was found that the best results were obtained when a grid was created where one grid cell equals 10cm on the ground. This allowed easy tracking of moving objects from humans to much larger objects. If tracking of smaller objects is required, higher resolution grid maps could be used. However, when higher resolution grids were used, the processing time was greater, and results were generally slightly poorer as noise became more prevalent.

After taking two occupancy grids, a difference map can be created by noticing the differences in the two scans. When generating a difference map, it is noticed that only cells that are occupied in the current grid map are of interest. Any other change in the grid map is either an object becoming occluded or the result of motion already detected from occupied cells. Because of this, the difference map is generated using Equation 1, given the current grid is G, the previous grid is G', and the difference map is D.

$$D_{i,j} = \begin{cases} 1 & \text{if } (G_{i,j} = 1) \wedge (G'_{i,j} = 0) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The output expected from this method can be seen in Figure 5 where the motion caused by the moving target has been captured.

6 Blob Segmentation

Once a difference map has been obtained, a simple two pass blob segmentation algorithm is applied to extract all 4-connected groups of cells. Because laser scan data only contains depth information, we cannot determine the centre of mass of detected objects which would provide the most accurate method for object tracking. Only the surface visible by the scanner, but we see that this provides information useful to tracking if the sur-

face of the object is tracked as though it was the entire target in question.

7 Hypothesis Creation

Hypothesis creation occurs by noticing changes in the blobs detected in the previous step. When a blob is seen in two separate locations, a movement hypothesis is created which consists of the position and velocity of the target. When creating hypotheses for objects moving in the scene, it is important to avoid creating hypotheses for every blob that can be seen. Not only will this create an excessive number of hypotheses, it will cause hypotheses to be created on top of already existing, managed hypotheses.

To solve this, each blob that is detected on each scan is determined to be a new blob, or one that is part of an existing hypothesis. Any new blob will cause hypotheses to be created between it and any nearby blob.

To determine if a blob is new, each blob is assigned a value B^{weight} where B^{weight} is in the range of 0 to 100. Any blob that is found to have B^{weight} above a certain threshold is regarded as new. The process used to calculate B^{weight} for each blob can be seen in Equation 2 where each blob coordinate is stored in B^{pos}, each hypothesis coordinate estimate is stored in list H^{est}.

$$B^{weight} = 100 - \sum_{hypos}^{i} \left[max(50 - dist(H_i^{est}, B^{pos}), 0) \right] \tag{2}$$

This simple approach results in a more manageable number of hypotheses, even when noise causes many blobs to be generated.

8 Hypothesis Management

Hypotheses are stored as positions of possible moving objects, as well as the object's estimated velocity and a weighting from 0 to 100. The weighting provides information on how likely the hypothesis is to be tracking a moving object against our constant velocity model. The velocity is maintained and projected using a Kalman Filter (see Section 9), and this projection is used to determine the hypothesis weight. Once a hypothesis is created, it will initially be given a low weighting to represent an initial uncertainty of the detected blob actually being a moving object. As each list of blobs is obtained for each frame, every hypothesis is checked to see if any of the given blobs could possibly be supporting a hypothesis by calculating the distance between blobs and the hypothesis estimate. If the distance is within a given threshold, the blob is deemed to be in support of the hypothesis and the weight of the hypothesis is increased based on this distance. The change in weight is determined by Equation 3 and relies on selected values of W^{inc} and W^{dec} which were experimentally determined to give best results at $W^{inc} = 60$ and $W^{dec} = 15$.

$$H^{weight} = H^{weight} + W^{inc} - W^{dec} * (dist(B^{pos}, H^{est})) \tag{3}$$

The process of weight modification can be seen in Figure 6, where hypotheses with data near their predictions are rewarded, and other hypotheses are discriminated.

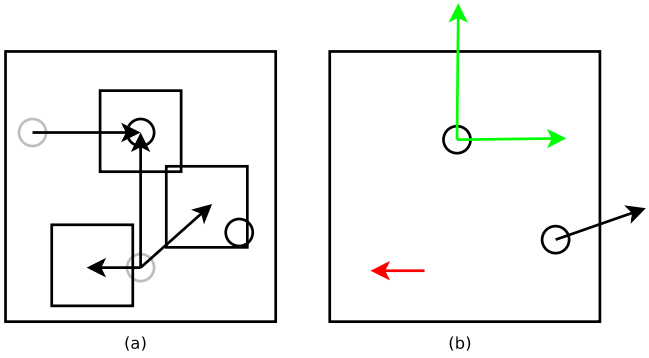

Figure 6: The process of tracking hypotheses. (a) Here each hypothesis (arrows) looks for data (circles) within their respective boxes. (b) Weightings are modified in reaction to the data found within their search areas. Green for a good match, red for no match, and black for a weak match.

The new detected blob is then passed into a Kalman filter, giving a new hypothesis to be used for next frame.

If multiple blobs are found to be near the estimate of a given hypothesis, the hypothesis is split, allowing the Kalman filter to track the two possible movements independently. This works particularly well when tracking people as the height of the scanner usually results in two legs being visible in the scene, with the dominant movement switching from leg to leg. By splitting the hypothesis, we can automatically switch the tracking, allowing us to always track both legs, and therefore track the person at a near constant velocity.

Any hypothesis that is left with no blobs near by, is updated along the same path but with a reduced weighting. This allows objects to become temporarily occluded, and yet still be tracked by the Kalman filter.

Hypotheses can also have their weight reduced when certain conditions hold, giving more control over the how objects are tracked. Any hypothesis that appears to be moving along a wall or any other structure has its weight halved as it is most likely that hypotheses are being generated due to noise. Once a hypothesis has a weighting that has reached a lower threshold, it is removed from the system and is no longer tracked.

The Kalman filter is designed to track parameters in the presence of noise, and because of this, it is noted that when initially tracking objects the Kalman filter will compensate for the lack of any pattern of constant velocity. As time goes on, however, we find that objects with constant velocity will be rewarded, while noise will not. This behavior of the Kalman filter forces us to ignore any hypothesis with insufficient data, and therefore moving objects are only reported as detected once at least m samples have been recorded. It was found that optimal tracking was achieved when $m = 4$.

An example of the hypotheses generated during operation can be seen in Figure 7. The lines represent hypotheses, where as the cross denotes that enough data has been obtained to trust the hypothesis as a moving object.

9 Kalman Filter

To keep track of the hypotheses, we have seen that a Kalman filter [Kalman, 1960] was employed, and parameters chosen such that objects moving at a constant velocity would be rewarded, and random noise would be discarded. The original Kalman filter equations can be seen below in Equations 4 to 8.

$$x_t^- = Ax_{t-1} + Bu_t \tag{4}$$

$$P_t^- = AP_{t-1}A^T + Q \tag{5}$$

$$x_t = x_t^- + K_t(z_t - Hx_t^-) \tag{6}$$

$$P_t = (I - K_tH)P_t^- \tag{7}$$

$$K_t = P_t^- H^T(HP_t^- H^T + R)^{-1} \tag{8}$$

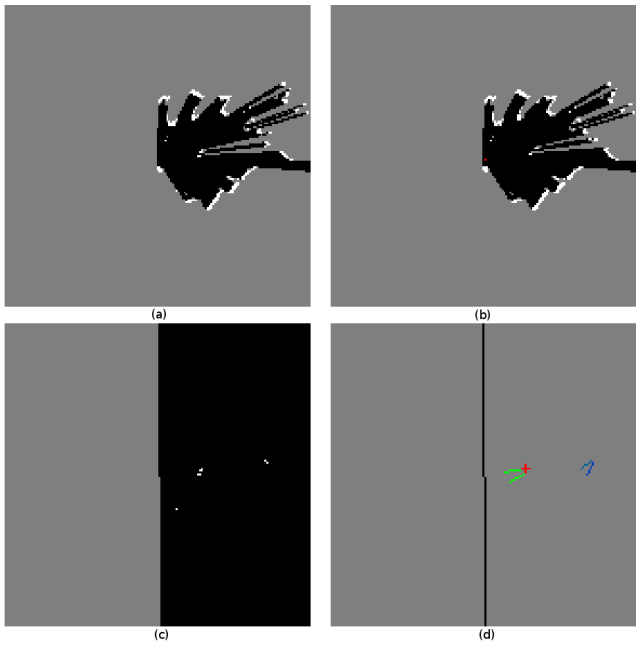

Figure 7: (a) Laser scan at time t (b) Laser scan at time $t-1$ (c) Difference Map with grey regions representing unknown space (d) Predicted velocities of moving objects

These equations allow for many different situations, but as we are aiming for a general approach to suit many types of moving objects, the Kalman filter was used constant velocity parameters resulting in a simplified set of equations seen in Equations 9 to 13.

$$x_t^- = x_{t-1} \tag{9}$$

$$P_t^- = P_{t-1} + Q \tag{10}$$

$$x_t = x_t^- + K_t(z_t - x_t^-) \tag{11}$$

$$P_t = (I - K_t)P_t^- \tag{12}$$

$$K_t = P_t^-(P_t^- + R)^{-1} \tag{13}$$

By using this constant velocity model, the only parameters left to define are the noise covariances Q and R. After experimenting with sample data, the following values for the covariance matrices were obtained in Equations 14 and 15 where units are in meters.

$$Q = \begin{bmatrix} 0.0005 & 0 \\ 0 & 0.0005 \end{bmatrix} \quad (14)$$

$$R = \begin{bmatrix} 0.006 & 0 \\ 0 & 0.006 \end{bmatrix} \quad (15)$$

This allows the Kalman filter to track objects, even when they change velocity quickly and continue on another constant velocity path. The weight of hypotheses depends on the error between the data and the Kalman estimate, so smaller errors will yield high probabilities for moving objects.

When tested on real data, the tracking performs well. An experiment was conducted where a mobile platform moved in front of a stationary laser scanner, and the odometry of the platform was recorded. Figure 8 shows the differences between the recorded odometry and the track produced by the Kalman filter used in this approach.

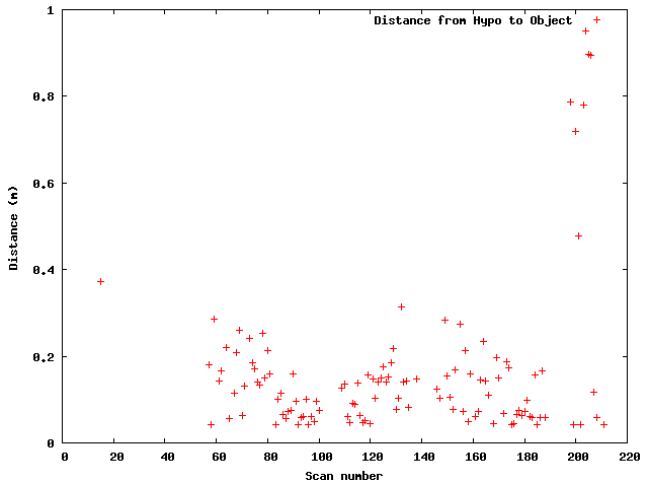

Figure 8: Distance between tracked hypothesis and actual target position. Estimate is usually within 0.2 meters of target.

10 Experimental Results

In order to determine the performance of the approach described, it was applied to a number of recorded scans with distinct properties to ensure that a high success rate was achieved in those cases.

10.1 Stationary Platform, Single Moving Object

In this experiment, the platform was placed in a cluttered room, with scans being obtained as in Figure 10. Over a 30 second period, 143 scans were taken where a target walked past the scanner a number of times, leaving the room for a short period. The human recorded track of the target can be seen in Figure 9, while the resulting track returned by the algorithm can be seen in in Figure 10.

The results obtained from this experiment can be seen in Table 1.

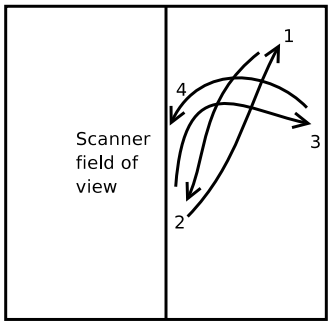

Figure 9: The track of a target in experiment 1. This track was recorded by a human operator. The target followed the course of the arrows in numerical order.

Table 1: Results from Stationary Platform, Single Moving Object Experiment

Experiment length	30 sec
Number of constant velocity paths taken	4
Number of constant velocity paths detected	4
Number of false detections	0
Number of velocity hypotheses made	38
Number of clearly incorrect velocity hypotheses made	0

The tracking was shown to work well at all the stages of the target movement (See Figure 10). We see that even though the targets tracked here are human legs and don't strictly follow a constant velocity model, they are still captured, and accurate velocity estimates obtained.

10.2 Moving Platform, No Moving Objects

Here the platform was moved around an office environment while the tracking process was performed to ensure that noise caused by the platform motion was ignored. We split the problem into two separate parts of platform translation and platform rotation, to see the performance in both cases.

Table 2 shows the number of false detections obtained while the platform is translating. We can see that performance is extremely good, with only 2 minor false detections.

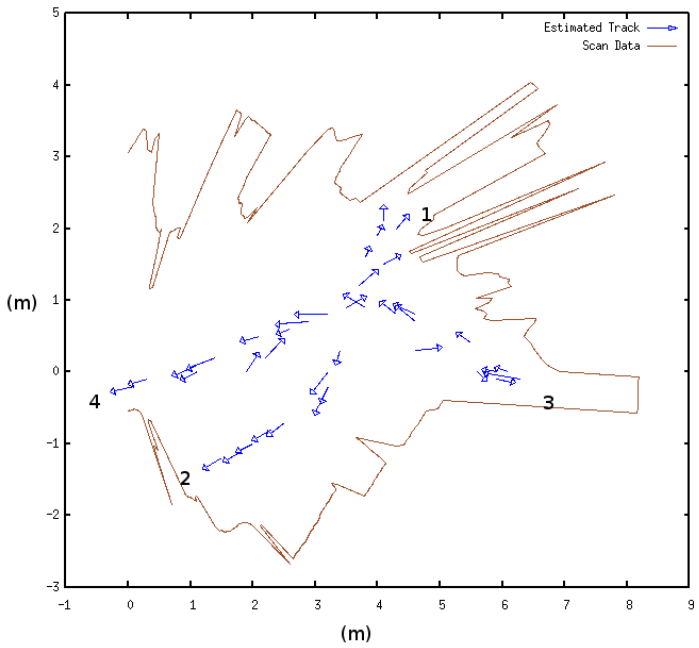

Figure 10: Results of tracking target in experiment 1. The vectors show the estimated velocity of the target at each point in time. We see relatively few hypotheses as a result of preferring to wait longer to be sure of a result.

Table 2: Results from no target, platform translation experiment

Experiment length	55 sec
Number of false detections	2
Duration of each false detection	1 frame

Table 3 shows the results from rotating the platform. We see that the number of false detections is greater than those obtained while translating, but as we expect platform translation to occur more frequently than platform rotation, this is acceptable.

Table 3: Results from no target, platform translation experiment

Experiment length	20 sec
Number of false detections	10
Maximum duration of any false detection	2 frames

10.3 Live Data

To ensure that the velocity data gathered from tracking is in fact valid, an experiment was conducted to use the information to gather further information about the tracked targets. An active head was used to direct cameras at the passing targets, and images were taken. The active head system contained significant delays, and as such, the cameras were aimed to lead the target by up to 3 seconds. This large leading required the data obtained from tracking to be accurate, and we see that this is the case.

The platform was placed at the end of a corridor, and passing targets were detected using the method described above before the cameras were trained on the target, and images captured. The results from this experiment can be seen in Table 4, and show that although the nearly every target was captured by the tracking system, 77% were captured by the cameras aimed at the targets. This figure is due to the delays in the active head system and were expected to be much worse. Any target remaining in the frame for a number of samples was captured, while targets quickly moving past the platform were missed as the cameras tried to catch up. Figures 11 to 14 show the images obtained from targets that were successfully captured by the system.

Table 4: Results from active head experiment

Number of passing targets	14
Number targets detected	13
Number of targets captured by camera	10
Percentage of targets captured	77%

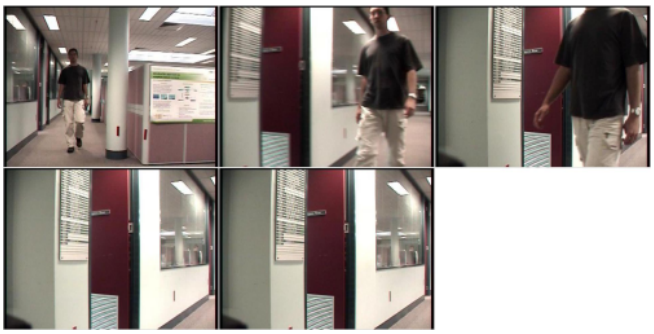

Figure 11: Individual target captured by the active head, targeting passing objects with an hypothesis method.

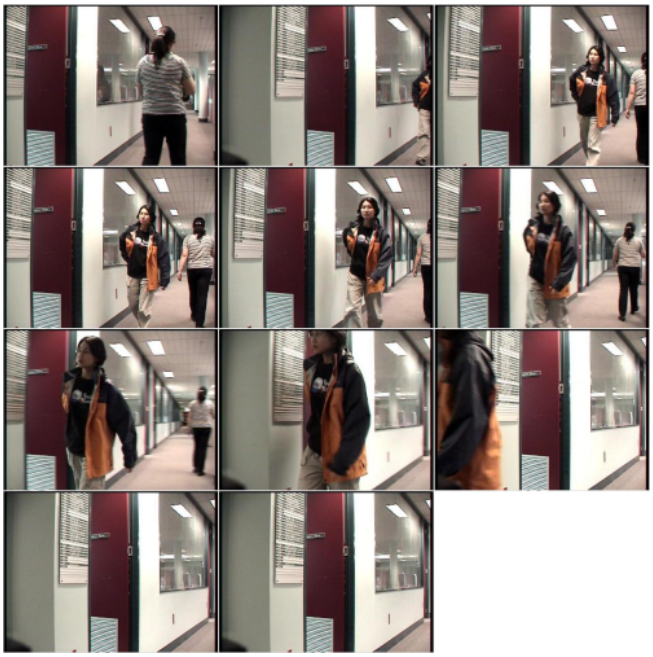

Figure 12: Multiple targets captured by the active head, targeting passing objects with an hypothesis method.

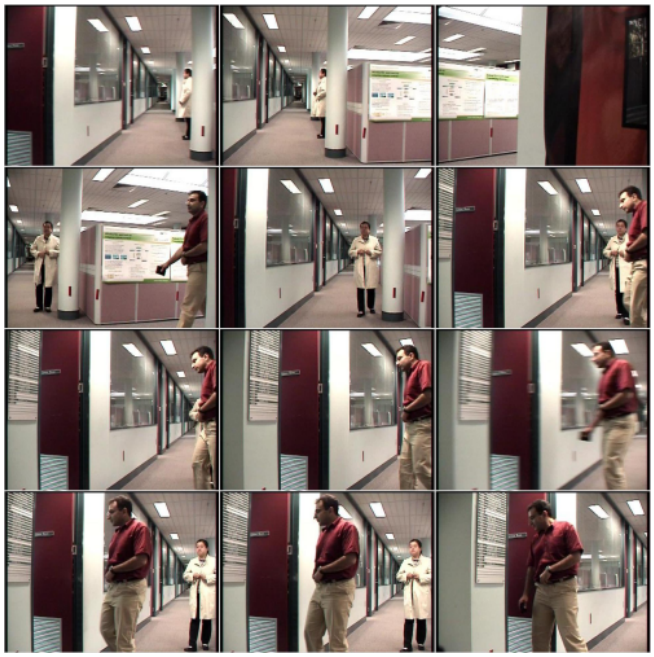

Figure 13: Multiple targets captured by the active head, targeting passing objects with an hypothesis method.

Figure 14: Multiple targets captured by the active head, targeting passing objects with an hypothesis method.

11 Conclusions

We see that by taking laser scan data, and combining scan matching, odometry, and hypothesis tracking techniques, we can successfully track objects moving within the scene without specific models of motion. We have also seen that when employed, it can easily provide velocity and position information that can be used in external systems for gathering further information. The generality of this approach makes it an appealing option for robots required to respond to a range moving targets in their environment.

Acknowledgements

This work was supported by funding from National ICT Australia. National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence program.

References

[Besl and McKay, 1992] P. Besl and N. McKay. A method for registration of 3–d shapes. 1992.

[Beymer and Konolige, 1999] D. Beymer and K. Konolige. Real-time tracking of multiple people using continous detection, 1999.

[Feyrer and Zell,] Stefan Feyrer and Andreas Zell. Detection, tracking, and pursuit of humans with an autonomous mobile robot.

[Fod *et al.*, 2002] Ajo Fod, Andrew Howard, and Maja J Mataric. Laser-based people tracking, 2002.

[Kalman, 1960] Emil Kalman, Rudolph. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[Lu and Milios, 1994] F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2d range scans. In *CVPR94*, pages 935–938, 1994.

[Needham and Boyle, 2000] Chris J. Needham and Roger D. Boyle. Tracking multiple sports players through occlusion, congestion and scale, 2000.

[Schulz *et al.*, 2000] Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. 2000.

[Siebel and Maybank, 2002] N.T. Siebel and S.J. Maybank. The application of colour filtering to real-time person tracking, 2002.

[Tao *et al.*, 1999] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. A sampling algorithm for tracking multiple objects. In *Workshop on Vision Algorithms*, pages 53–68, 1999.