Proceedings of the 1999 IEEE
International Conference on Robotics & Automation
Detroit, Michigan • May 1999

Continuous Mobile Robot Localization: Vision vs. Laser

J. A. Pérez, J. A. Castellanos, J. M. M. Montiel, J. Neira, J. D. Tardós
Universidad de Zaragoza
Departamento de Informática e Ingeniería de Sistemas
c/María de Luna, 3, E-50015 Zaragoza, SPAIN
email: {jperez,jacaste,josemari,jneira,tardos}@posta.unizar.es

Abstract

In this work we describe a comparative study of the performance of map-based robot localisation processes based on diverse sensing devices such as monocular and trinocular vision systems, and laser rangefinders. We study both the precision (error with respect to the true values) and robustness (sensor measurements correctly paired with map features) of each localisation process. The experiment design that we have followed allows us to compare these processes under exactly the same conditions. We conclude that comparable precision levels can be attained with each of the three sensors. With respect to robustness, monocular and trinocular vision pose more complex matching problems than laser, requiring more elaborate solutions to make the process robust.

Keywords

Robot Location, Monocular Vision, Trinocular Vision, Laser Rangefinder

1 Introduction

The problem of precisely and robustly locating a mobile robot following a trajectory in a known environment has been treated extensively in the literature [1, 2, 3, 4]. In general, the basic idea is always the same: at each location of the robot trajectory the localisation system perceives the environment with some sensor, compares the sensorial information with the expected values (predicted from an *a priori* map), and corrects the available robot location, generally given by an odometric system, to make the perceived data match better with the expected data. The precision and robustness of this process depends on several factors, being the most significant: the sensor used, the available map, and the matching process. Some researchers have reported studies of the precision of localisation processes using some sensors. However, these works have not established what the *true* error is, because the true location of the robot is not known (only error bounds are obtained).

We have studied this problem to try to answer the following questions: under exactly the same conditions, which sensor performs better in carrying out the localisation process? Can we attain with monocular vision comparable precision levels than with laser? Or is stereo vision (in this case, trinocular vision) necessary to attain sufficient precision levels? In order to adequately answer these questions, the following must be taken into account:

1. In order to be able to measure and compare the precision of the results in each case, we need to know the *true* robot location at each step of the trajectory, and have a precise description of the environment.

2. We must be able to test the location processes using the same *a priori* information and for the same trajectory. We may even want to try out different matching processes for the same sensor data, to compare robustness.

We have designed an experiment that allows us to make this comparative study guaranteeing that the former requirements are met. Due to their popularity, we have concentrated in comparing laser rangefinders and vision systems, monocular and trinocular. Vision is a more versatile sensor, so we are basically interested in knowing whether it can achieve comparable precision and robustness levels. We have included both monocular and trinocular vision to see whether only one camera, where there is no depth information, may attain satisfactory precision and robustness levels. We use trinocular instead of stereo vision because the redundancy given by the third camera makes the correspondence problem simpler.

This paper is organised as follows: in section 2 we describe the experiment design. Section 3 contains a description of the location process carried out for each sensor. The obtained results are discussed in section 4. Finally, in section 5 we draw the main conclusions of this work. In order to have sufficient space for presenting our results and discussing them, we have avoided including any mathematical detail related to the representation and fusion of sensorial information in the localisation processes. All mathematical details of these processes have been described extensively and can be found in [4, 5].

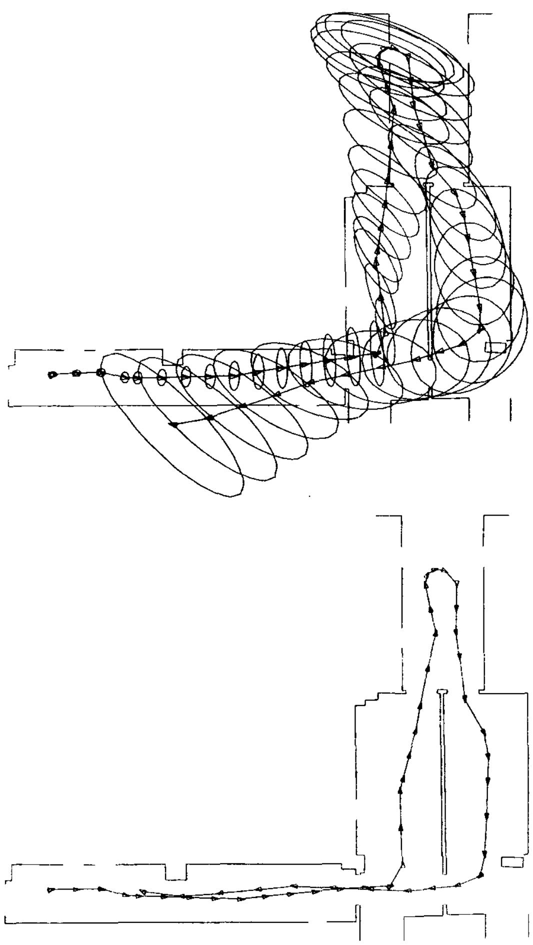

Figure 1: Robot location according to odometry (top) and the theodolites (bottom).

2 The Experiment

The experiment was carried out using the following equipment: a Labmate mobile robot, a laser range system and a trinocular vision system, both mounted on the mobile robot, and a pair of theodolites, used as a precise location measurement equipment, independent from the sensors and the mobile robot.

The *environment model* is a set of vertical edges, corresponding to wall corners and door frames, whose location was measured with the theodolites. The location of vertical walls was calculated using this information. The resulting 2D environment map is shown in fig. 1.

The robot was programmed to follow a trajectory in our laboratory. At each step of the path, the robot location according both to odometry, and measured with the theodolites, was obtained. Being the most precise measurement independent from sensor observations, it is considered the *true* robot location. In fig. 1, both the odometric (with uncertainty ellipsoids) and the theodolite robot location are shown, highlighting the cumulative nature of odometric errors that eventually make the measurement useless.

At each step of the robot trajectory, the environment was sensed using both trinocular vision (fig. 2), and the laser rangefinder. Fig. 3 shows the environment information obtained at step 22 of the robot trajectory. Large vertical edges (more than 100 pixels) are extracted from the three images: the ones corresponding to the central image are considered monocular infinite projection rays of corners and door frames, and the rays of the three images are matched to obtain the location of trinocular vertical edges also corresponding to corners and door frames [7]. The laser scan is segmented to obtain segments corresponding to walls [5]. In the next section we describe the three experiments that we carried out with this information. A detailed description of this experiment, including complete robot and sensor specifications, as well as all data sets, can be found in [6].

3 Robot Localisation

The process of refining the estimated location of the mobile robot given a sensor measurement E of a certain environment feature M, which is applicable to the three types of sensor information that we consider (monocular vision rays, trinocular vision points and laser segments), is carried out in the following way:

1. The imprecision in both the estimated robot lo-

Figure 2: Trinocular Image Set at step 22 of the robot trajectory.

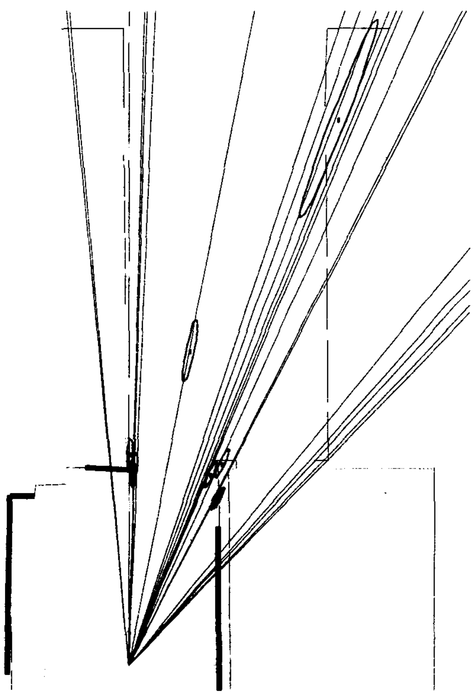

Figure 3: Features observed at step 22 of the robot trajectory : laser segments (thick segments) corresponding to walls, monocular rays (infinite projection rays) corresponding to corners and door frames, and trinocular points (thick points) corresponding also to corners and door frames.

cation and the sensor measurement is represented using a probabilistic model, the SPmodel [8].

2. At each step in the robot trajectory, the sensor acquires a set of measurements of the environment features. A matching process must decide either to pair each of these measurement with some map feature, or to label the measurement as spurious. This matching process is different for each type of sensor, and it is explained in each case below.

3. The possible correspondence between a certain sensor observation and a map feature is verified using a statistical test (the chi-square test) based on the squared Mahalanobis distance between the observation and the feature.

4. Having found that sensor observation E is compatible with map feature M, we establish a non-linear measurement equation that states that the location of E and M must coincide, and fuse this information using a specialised version of the Kalman Filter, the SPfilter [4], to refine the estimated robot location.

3.1 Localisation using Monocular Vision

As it can be seen in fig. 3, the matching process between projection rays given by monocular vision and vertical edges corresponding to corners and door frames is very complex, due to several facts:

- There are many environment visual features which are not represented in the map.
- Some corners represented in the map cannot be detected by the vision system due to illumination conditions.
- The projection ray is infinite, so there is no information about the distance to the observed feature.

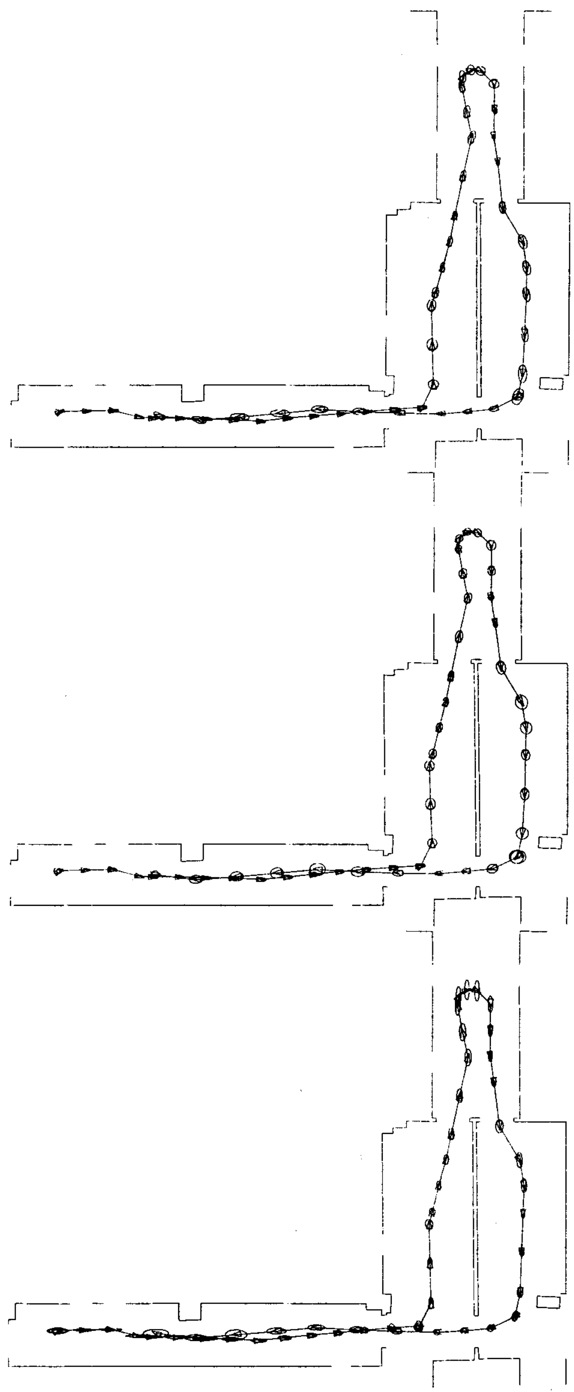

Figure 4: Estimated trajectory using Monocular Vision (top), Trinocular Vision (middle) and Laser (bottom). Uncertainty × 3.

- The proximity of different points of the map and the uncertainty in the initial odometry estimation of the robot position causes that multiple map corners match as candidates for pairing with an observation.

To minimise the risk of the robot getting lost, secure pairings (projection rays that have a unique map corner as pairing candidate) are integrated first. Next, if there are non-secure pairings, a limited number of them is integrated. The trajectory resulting from this process is shown in 4, top.

3.2 Localisation using Trinocular Vision

The trinocular vision system gives us the location of vertical edges in 3D space (which correspond to points in 2D space). These points are obtained by a correspondence algorithm [7]. At each point of the trajectory, the predicted location of potentially visible corners is calculated, and again the chi-square compatibility test is applied to each of the trinocular points with each of the predicted corners. If more than one corner matches a trinocular point, the corner with the smallest square Mahalanobis distance is selected. Because the correspondence algorithm has eliminated most of the spurious information, this process is much simpler than the one of monocular vision, although the depth precision of trinocular vision is rather low. The resulting trajectory is shown in 4, middle.

3.3 Localisation using Laser

Laser segments are obtained by application of a segmentation technique [5] to the set of data gathered by the laser rangefinder. The matching process is quite simple. Again, at each point of the trajectory, the predicted location of potentially visible walls is calculated, and a suitable pairing for each laser segment is obtained by applying the chi-square compatibility test with each of the predicted walls. The segment is paired to the wall that satisfies the test, and whose extension includes the segment location. If there is more than one, the wall with the smallest Mahalanobis distance is selected. The trajectory resulting from applying this process is shown in 4, bottom.

4 Discussion

In this section we compare the results obtained in the localisation process using the three sensor systems, analysing the precision, robustness and practicality of each approach.

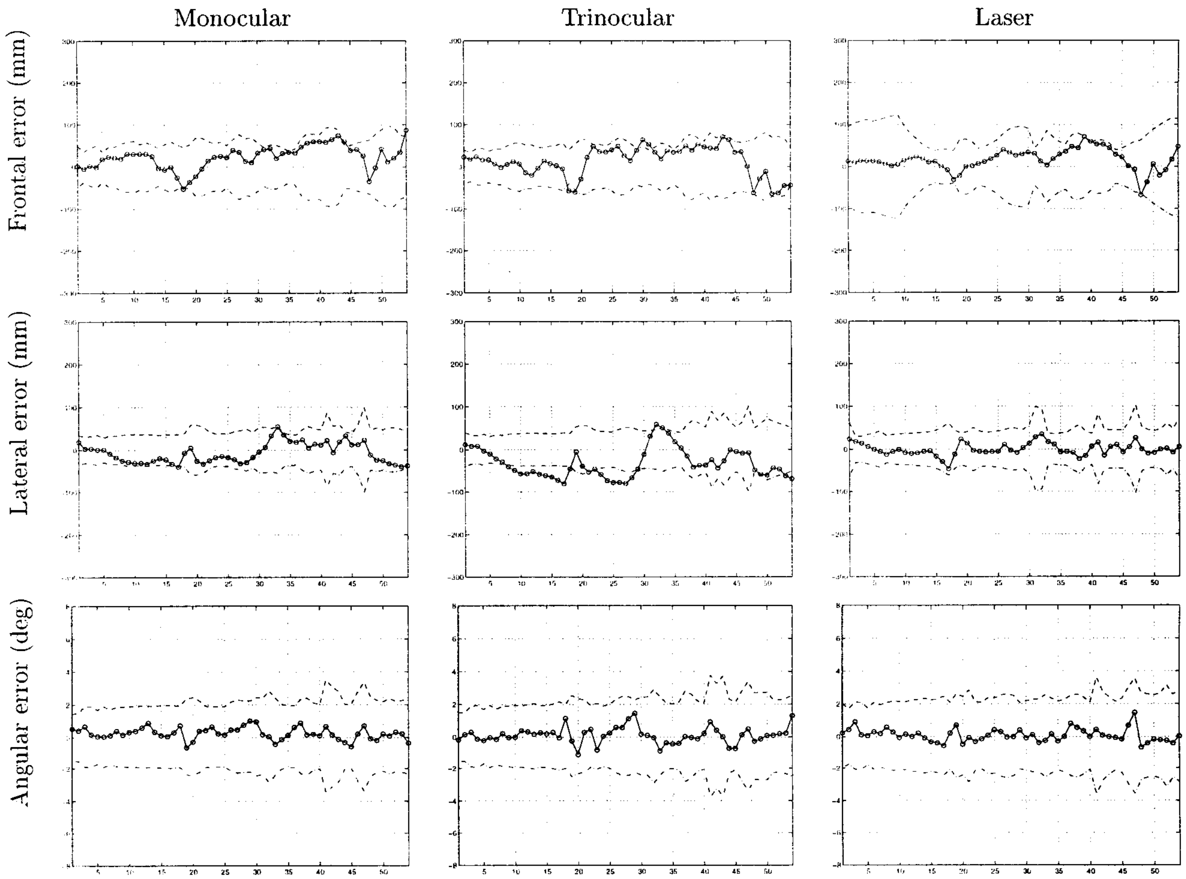

Figure 5: Real errors and 95% uncertainty interval during the localization process for each sensor.

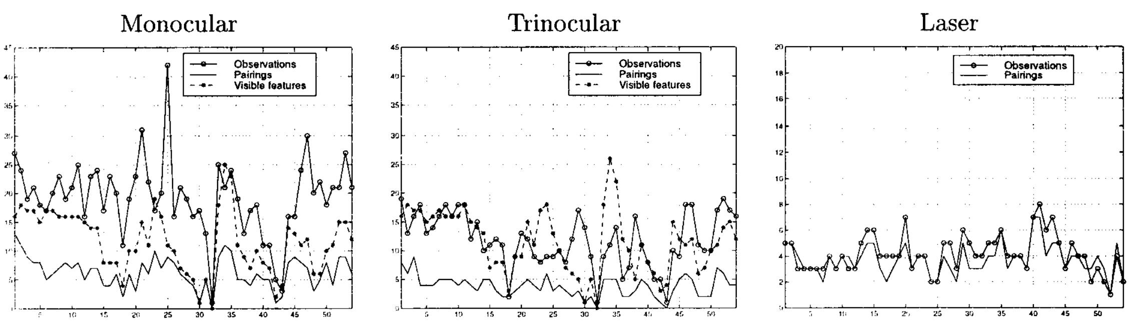

Figure 6: Number of observations and pairings with Monocular Vision (left), Trinocular Vision (middle) and Laser (right).

4.1 Precision

Figure 5 shows the errors obtained in the robot location estimation relative to the *true* location, measured with the theodolites, and the computed location uncertainty, drawn as the 95% confidence interval. A summary of the precision obtained with each sensor is given in Table 1. All three sensors allow to estimate the robot location with a worst-case position error of

	Frontal (mm)		Lateral (mm)		Angular (deg)	
	avg	max	avg	max	avg	max
Mono.	30	87	21	55	0.4	1.0
Trino.	32	71	41	81	0.4	1.5
Laser	23	71	11	46	0.3	1.5

Table 1: Average and worst case localization errors

	Acq.	Proc.	Match. and Loc.	Tot.
Mono.	0.04	0.15	0.03	0.22
Trino.	0.04	0.55	0.02	0.61
Laser	0.10	0.06	0.01	0.17

Table 2: Processing times for each sensor measured in seconds on a SPARCstation 20

less than 10cm, and an orientation error of less than 2deg. This precision can be adequate for successful autonomous navigation in most environments. The computed uncertainty gives a realistic measurement of the true errors obtained in the localisation process, allowing the system to stop the robot and launch a recovery procedure in the event that the robot gets lost.

The lateral precision obtained with trinocular vision is somewhat lower than with monocular vision, mainly due to the fact that a given feature must appear in the three images in order to be located by the trinocular vision systems, and thus in some cases less features may be fused. Precision can be increased using a longer baseline (it was 40cm in our experiment), at the cost of a small loss in field of view, and a bulkier installation.

It is important to highlight that the resulting error plots in the three cases are very similar. That is, the three sensors are coherent in their estimations of the robot location. This shows the validity of the approach based on extracting and fusing geometrical information from the sensors, and supports the plausibility of a multisensor fusion approach based on geometric information.

4.2 Robustness

The robustness of the localisation process is mainly related with the reliability of the matching between observations and map features: incorrect pairings can make the estimation converge to an incorrect solution with covariances under-estimating the real errors obtained. This forces us to adopt a conservative approach: it is better to reject a correct matching than risking to accept an incorrect one. In this work, we have used the chi-square test based on the square Mahalanobis distance as the basic matching tool, with the following results:

- In the case of monocular vision, the lack of depth information produces many potential matches for a given observation. This leads easily spurious pairings. In order to limit this possibility, observations with one candidate map feature are fused first, and the number of observations with several candidate map features that are fused next is limited. This makes the number of accepted pairings somewhat low in some cases (see fig. 6, left).
- In the case of trinocular vision, the redundancy provided by the third camera allows to successfully reject nearly all incorrect pairings between vertical edges in each image. Additionally, the availability of depth information reduces the number of potential pairings between trinocular points and map corners and door frames, giving as result a much more robust procedure. There are however many environment features observed by the trinocular system but not included in the map, and thus the number of pairings is also low in some cases (see fig. 6, center).
- The information provided by the laser rangefinder allows to locate with precision the walls around the robot, most of which can be matched in without errors (fig. 6, right).

4.3 Practicality

Taking into account processing time, laser is superior (Table 2). However, given the simple image processing techniques that we have used (only long vertical edges are extracted), both monocular and trinocular vision can be used to localise in real time a robot moving at 1m/s, with precision less than 10cm.

In the case of laser, all range information is obtained in one horizontal scan, which can constitute a problem if there happens to be many obstacles at this particular working height. If the scan were performed with the robot in motion, either an additional error would be introduced, making the results less precise, or additional processing would be required to compensate for this distortion. Apart from that, the information provided is simple, robust, and easy to process. By contrast, visual information is much richer and versatile, at the cost of more processing time, and more difficulties in finding correct pairings.

Another difficulty for the practical use of vision is the criticality of the calibration procedure, mainly in the case of trinocular. In our experiment the system was calibrated using a pattern close to the robot, resulting in a poor precision at middle and long distances. On the other hand, the higher cost, size and weight of laser rangefinders make them less suitable for some applications like autonomous wheeled chair navigation.

5 Conclusions

Map-based robot localisation based on vision, both monocular and trinocular, achieve levels of precision equivalent to those attained by the use of laser sensors. Laser is computationally simpler, and it gives precise and robust results. It is however more limited, in the sense that no additional information may be obtained from a scan. Since the main drawback of using vision has been robustness, the development of robust matching techniques for vision is one of the subjects of future work. Another important factor affecting the precision and robustness of vision is the fact that the *a priori* map contains information that includes many features detectable with laser (walls), but few visual features (only corners and door frames). It is costly and difficult to obtain a precise and complete *a priori* map. This suggests another line of future work: using the sensor to locate the robot and build a map simultaneously.

References

[1] J.J. Leonard and H.F. Durrant-Whyte, *Directed Sonar Sensing for Mobile Robot Navigation*, Kluwer Academic Publishers, London, 1992.

[2] X. Lebègue and J.K. Aggarwal, "Significant line segments for an indoor mobile robot," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 6, pp. 801–815, 1993.

[3] J. Horn and G. Schmidt, "Continuous localization of a mobile robot based on 3d-laser-range-data, predicted sensor images and dead-reckoning," *Journal of Robotics and Autonomous Systems, Special Issue "Research on Autonomous Mobile Robots"*, vol. 14, pp. 99–118, 1995.

[4] J. Neira, J. Horn, J.D. Tardós, and G. Schmidt, "Multisensor mobile robot localization," in *Proc. 1996 IEEE Int. Conf. on Robotics and Automation*, Minneapolis, USA, 1996, pp. 673–679.

[5] J.A. Castellanos and J.D. Tardós, "Laser-based segmentation and localization for a mobile robot," in *Robotics and Manufacturing vol. 6 - Procs. of the 6th Int. Symposium (ISRAM)*, F. Pin, M. Jamshidi, and P. Dauchez, Eds., pp. 101–108. ASME Press, New York, NY, 1996.

[6] J.A. Castellanos, J. M. Martínez, J. Neira, and J.D. Tardós, "Testbed for mobile robot navigation: The CPS Experiment," Technical Report RR-97-11, CPS, Zaragoza, Spain, 1997.

[7] J.M Martínez, Z. Zhang, and L. Montano, "Segment-based structure from an imprecisely located moving camera," in *IEEE Int. Symposium on Computer Vision*, Florida, November 1995, pp. 182–187.

[8] J.D. Tardós, "Representing partial and uncertain sensorial information using the theory of symmetries," in *Proc. 1992 IEEE Int. Conf. on Robotics and Automation*, Nice, France, 1992, pp. 1799–1804.