Huang, Henry and Maire, Frederic D. and Keeratipranon, Narongdech (2007) Bearing-Only SLAM with Indistinguishable Landmarks. In Proceedings The 2007 Australasian Conference on Robotics and Automation (ACRA 2007), Brisbane, Australia.

Bearing-Only SLAM with Indistinguishable Landmarks

Henry Huang, Frederic Maire and **Narongdech Keeratipranon**
Faculty of Information Technology
Queensland University of Technology
Box 2434, Brisbane Q 4001, Australia
Email: {h5.huang@student., f.maire@, n.keeratipranon@student.}qut.edu.au

Abstract

To navigate successfully, a mobile robot must be able to estimate the spatial relationships of the objects of interest in its environment. The main advantage of a bearing-only SLAM system is that it requires only a cheap vision sensor for navigation. In this paper, we present a bearing-only SLAM method for a 2D environment with visually indistinguishable landmarks (*Unknown Data Association*). When the frame rate of the vision sensor is high, it is easy to track the landmark correspondences. When the frame rate is low, we estimate the landmark positions by computing the most likely hypothesis based on bearing measurements. A bearing-only SLAM system that employs this method does not require any other sensors such as range sensors or wheel encoders. Due to its low cost, this system is ideally suited for domestic robots such as autonomous lawnmowers and vacuum cleaners.

1 Introduction

What is a *Map*? For a mobile robot, any spatial relationship between the objects of interest in its environment is a map. One of the fundamental challenges in mobile robotics is the *Simultaneous Localization and Mapping* (SLAM) problem. A SLAM system builds incrementally a map of an unknown environment from observations made by a robot.

There exist well developed solutions to the SLAM problem for the case where the mobile robot is equipped with a sensor that provides both range and bearing measurements to landmarks [Leonard and Durrant-Whyte, 1991; Spero, 2005; Zunino and Christensen, 2001]. However, the sensors these solutions rely on are considered too expensive for domestic robots. Thanks to recent advances in computer vision and cheaper cameras, vision sensors have become popular for SLAM [Bailey, 2003; Costa *et al.*, 2004; Davison, 2003; Davison *et al.*, 2004].

One of the fundamental tasks of a SLAM system is the estimation of the landmark positions in an unknown environment. This task is called *Landmark Initialization*. In SLAM, the *Data Association* problem, also known as the *Correspondence* problem, is the problem of relating sensor measurements to features in the environment that is being explored. The uncertainty of landmark correspondence occurs due to the robot's inability to properly match the views of the same landmark from different locations. Therefore there exists the possibility that the robot wrongly associates landmarks. This problem is generally referred to as the *Unknown Data Association* problem or the *Unknown Correspondence* problem [Spero, 2005].

Costa *et al.* [Costa *et al.*, 2004] presented an iterative method using only bearing measurements to solve the landmark initialization problem with unknown data association. Their method requires a time lag to wait for sufficient observations. Spero [Spero, 2005] proposed a multiple-hypothesis algorithm to the SLAM problem with unknown data association. A graph matching approach was presented in [Spero, 2005] to compare the similarities of the subgraphs extracted from each hypothesis. This algorithm was based on range and bearing measurements from a laser range finder. It is desirable to solve the unknown data association problem in SLAM when only bearing measurements are given.

Some recent solutions employ *Structure From Motion* (SFM) to solve the bearing-only SLAM problem [Goncavles *et al.*, April 2005; Jensfelt *et al.*, 2006; Huang *et al.*, 2007]. The common approach for SFM is to estimate object positions by triangulation using the bearing measurements. Huang *et al.* [Huang *et al.*, 2007] solved the landmark initialization problem of bearing-only SLAM with the bearings extracted from only two observations along a linear trajectory. However, the authors assumed that the landmarks were distinguishable from different observation points. Since all the images between these two observation points were ignored, the data association might not be maintainable if the landmarks were visually indistinguishable.

In this paper, we extend the work of [Huang *et al.*, 2007] and propose a method for bearing-only SLAM for a 2D environment with indistinguishable landmarks. When the frame rate of the vision sensor is high, it is easy to track the land-

mark correspondences. The apparent angles and the visual sequences of the landmarks are employed for the tracking (will be fully described in Section 2). When the landmark correspondences are not trackable due to a too low frame rate, our method computes the most likely hypothesis of the relative landmark positions. The computation requires only the landmark bearings derived from three panoramic views made in a linear trajectory.

Next section describes how to track the landmark correspondences. Section 3 presents our method to solve the bearing-only SLAM problem when the landmark correspondences are unknown. In Section 4, we present some experimental results using a car-like robot in an outdoor environment. Section 5 discusses future work and concludes this paper.

2 Tracking the Landmark Correspondence

In this section, we introduce our approach for tracking the landmark correspondences when the frame rate of the vision sensor is high. Once the landmark correspondences are available (the data are associated), the SLAM problem can thus be solved using existing methods (for example, [Huang *et al.*, 2005; Jensfelt *et al.*, 2006]).

To illustrate our approach, consider a simple environment (Fig. 1) with four indistinguishable landmarks $\{L_1, ... L_4\}$. Assume that the robot measures the bearings of the landmarks in counter-clockwise direction. We define the *Visual Order* of a set of landmarks as the perception sequence of the landmarks. For instance, the visual order at observation point O_1 with respect to the four landmarks is (L_3, L_4, L_1, L_2).

In Fig. 1, the bearing of the jth landmark in the visual order at the observation point O_i is denoted by β_i^j. The environment is divided into three zones: zone A is the convex hull formed by the four landmarks; the hashed regions make up zone C and the rest belongs to zone B. If a straight trajectory of the robot is within the union of zone A and zone B, then the visual order stays the same along the trajectory. Figure 1 shows that the visual orders at both O_1 and O_2 are (L_3, L_4, L_1, L_2). The corresponding sequences of bearings at O_1 and O_2 are $(\beta_1^1, \beta_1^2, \beta_1^3, \beta_1^4)$ and $(\beta_2^1, \beta_2^2, \beta_2^3, \beta_2^4)$ respectively. The landmark correspondences at O_1 and O_2 can be found easily. Hence, given the locations of O_1 and O_2, the bearings β_1^1 and β_2^1 are sufficient to determine geometrically the position of L_3. Similarly, the positions of L_4, L_1 and L_2 can be determined. The visual order will be altered if the robot moves from zone B to zone C. We will show how to determine the situation when the robot moves across the boundary between zone B and zone C.

At any observation point, the *Apparent Angle* between two consecutive landmarks with respect to the visual order is the difference in the bearing measurements of these two landmarks. If an observation point is inside zone A, such as O_2 in Fig. 1, all the apparent angles will be smaller than 180 degrees. At an observation point in zone B, such as O_1 in Fig.
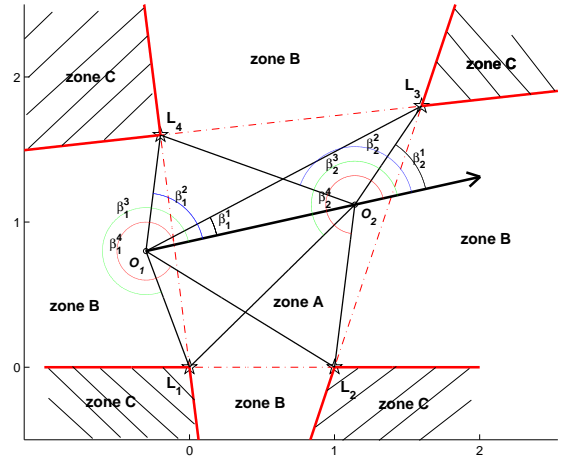

Figure 1: If a straight trajectory is within the union of zone A and zone B, then the visual order at any point along the trajectory remains unchanged. In this example, the landmark correspondences at O_1 and O_2 are easy to solve.

1, the robot will observe an apparent angle greater than 180 degrees. If the robot's location is on the boundary between zone B and zone C, in addition to an apparent angle greater than 180 degrees, one of the apparent angle becomes 0.

Hence, when the frame rate is high, landmark correspondences at two observation points along a straight trajectory can be derived from the visual orders by tracking the evolution of apparent angles. If there exists a point along the trajectory where an apparent angle is greater than 180 degrees and another apparent angle is 0, then the visual order after the robot crosses this point will be altered. The alteration is only with respect to the two landmarks whose apparent angle is 0, the indices of these two landmarks in the visual order are swapped.

3 Computing the Most Likely Landmark Positions

This section describes our bearing-only SLAM method with indistinguishable landmarks. We assume the landmark correspondences are not trackable from one image to the subsequent image due to the low frame rate or very fast robot's motion. For example, between two consecutive images are taken, the robot may have moved more than one meter. We compute the most likely landmark positions by comparing the geometric similarity of all the possible hypotheses of the relative landmark positions derived from the observations.

One fundamental problem facing a robot-based mapping system is how to translate the relative positions of objects with respect to the moving frame of the robot (as well as the position uncertainties) into relative positions with respect to a static frame in the environment. In [Huang *et al.*, 2005], we introduced an *Uncertainty Analysis Method* (UAM) to solve

this problem. The moving frame attached to the robot, called the *robot-based frame*, is denoted by $\mathcal{F}_\mathcal{R}$. The static frame attached to the landmarks, called the *landmark-based frame*, is denoted by $\mathcal{F}_\mathcal{L}$. In this paper, to compare the geometric similarity we employ UAM to transform the spatial uncertainties of the landmarks from $\mathcal{F}_\mathcal{R}$ to $\mathcal{F}_\mathcal{L}$. Since the robot is moving, the base of $\mathcal{F}_\mathcal{R}$ is changing too. It is essential to compute the most likely landmark positions with respect to the frame $\mathcal{F}_\mathcal{L}$ attached to two fixed landmarks.

Let β denote a landmark bearing measured at an observation point O. Assume that the vision error for the bearing is smaller than ϵ, the landmark position is contained in the *vision cone* formed by the two rays $\beta - \epsilon$ and $\beta + \epsilon$ rooted at O. Consider an environment that contains n landmarks visually indistinguishable, there are n vision cones rooted at O. Each vision cone is centered in the bearing direction. When two observations are made along a straight trajectory, there exist at most n^2 polygons generated by intersecting all the vision cones. Figure 2 shows an environment with three landmarks ($n = 3$). When two observations are made at O_1 and O_2, nine intersection polygons are generated. Because each vision cone must contain a landmark, there are six ($n!$) different possible configurations of the three landmarks as displayed with solid lines in Fig. 3.
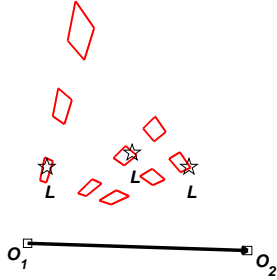

Figure 2: An environment that contains three indistinguishable landmarks. Nine landmark polygons are generated by intersecting all the vision cones rooted at O_1 and O_2.

To determine the most likely landmark positions from all the hypotheses, more observations are needed. Figure 4a shows a general situation. The robot moves further and two more observations are made at O_3 and O_4 in another straight trajectory. Figure 4b shows all the polygons generated from these two pairs of observations. We will illustrate our method for finding out the most likely landmark positions.

Let $\mathcal{H}_{o_1 o_2}$ and $\mathcal{H}_{o_3 o_4}$ be the sets of possible hypotheses of the landmark polygons derived from $O_1 O_2$ and $O_3 O_4$ respectively, such that $\mathcal{H}_{o_1 o_2} = \{h_i \mid i \in 1...6\}$ and $\mathcal{H}_{o_3 o_4} = \{k_j \mid j \in 1...6\}$. The most likely hypothesis is determined by comparing the geometric similarity between h_i and k_j.

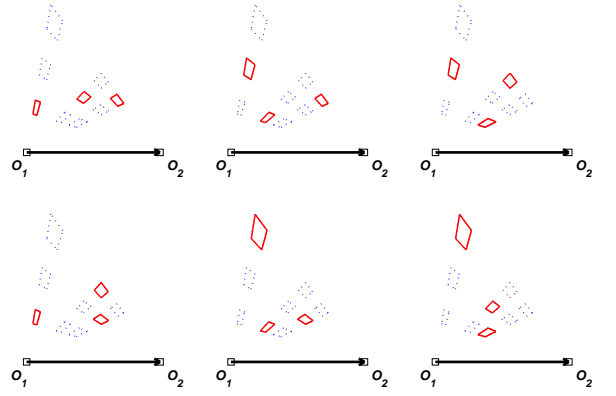Both h_i and k_j contain three polygons which represent the

Figure 3: Six possible hypotheses derived from nine landmark polygons satisfying the constraint that each vision cone must contain one landmark. Each hypothesis contains three landmark polygons.
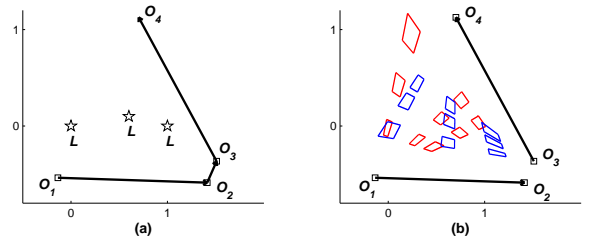

Figure 4: (a) A configuration with two pairs of observations of three indistinguishable landmarks. (b) All the landmark polygons derived from the two pairs of observations, $O_1 O_2$ and $O_3 O_4$.

uncertainties of the landmark positions. However, h_i and k_j are not expressed in the same frame; h_i is expressed in the frame $\mathcal{F}_\mathcal{R}$ attached to O_1, O_2 and k_j is expressed in $\mathcal{F}_\mathcal{R}$ attached to O_3, O_4. For the comparison of the geometric similarity between h_i and k_j, we employ the UAM to transform h_i and k_j to the same landmark-based frame $\mathcal{F}_\mathcal{L}$ attached to two landmarks. In other words, the uncertainties of the two landmarks are squeezed to zero (without uncertainty) and all the uncertainties of these two landmarks are transferred to the third landmark. The geometric similarity of h_i and k_j can be obtained from the intersection ratio between the third landmark polygons of h_i and k_j expressed in $\mathcal{F}_\mathcal{L}$.

Figure 5 illustrates the computation of the geometric similarity between h_i and k_j. Figure 5a shows h_i expressed in the $\mathcal{F}_\mathcal{R}$ attached to O_1 and O_2. Figure 5b shows k_j expressed in the $\mathcal{F}_\mathcal{R}$ attached to O_3 and O_4. After the transformation using the UAM, h_i and k_j expressed in $\mathcal{F}_\mathcal{L}$ are shown in Fig. 5c and 5d respectively. By construction, the two landmarks denoted by stars at coordinates $[0\ 0]'$ and $[1\ 0]'$ are certain. The uncertainties of the two landmarks are transferred to the third landmark. Figures 5c and 5d are unifiable since they are
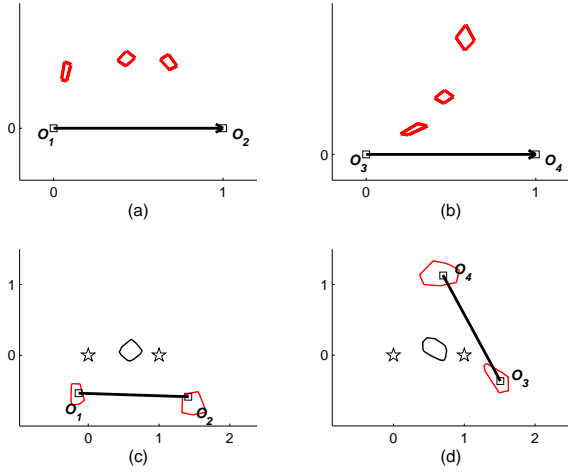
Figure 5: (a) and (b) show two hypotheses h_i and k_j in $\mathcal{F}_\mathcal{R}$. The spatial relationships and uncertainties of all objects in the landmark-based frame $\mathcal{F}_\mathcal{L}$ are shown in (c) and (d).

with respect to the same frame $\mathcal{F}_\mathcal{L}$ attached to the two landmarks denoted by stars. The geometric similarity of h_i and k_j can be computed from the intersection ratio of the third landmarks of h_i and k_j expressed in $\mathcal{F}_\mathcal{L}$. The locations of the observation points can be estimated using the same transformation method simultaneously. Figure 6 shows the estimated spatial relationships of landmarks and observation points after the unification of Figs. 5c and 5d.
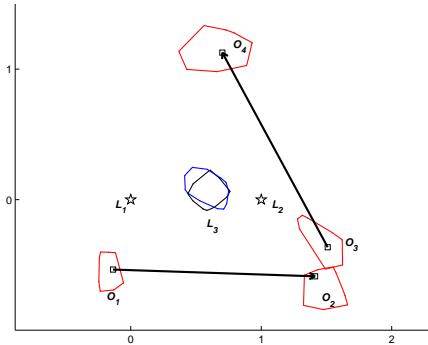

Figure 6: By computing the most likely hypothesis, the spatial relationships and the uncertainties of landmarks and robot's locations can be estimated simultaneously.

Our method computes all the possible hypotheses of landmark polygons derived from two pairs of observations. With n landmarks, each pair of observation contains $n!$ hypotheses, the computational complexity is $O((n!)^2)$. Without the knowledge of data association, all the permutations of landmarks in each hypothesis have to be computed. The total complexity becomes $O((n!)^3)$. When n is large, the com-
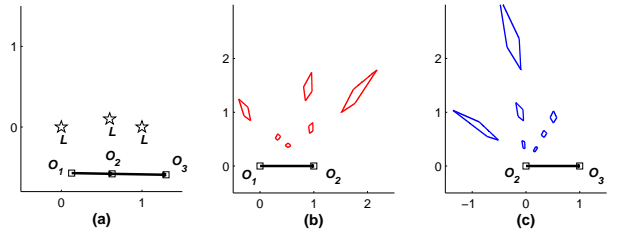

Figure 7: (a) Three observations are made along a straight trajectory with three landmarks. (b) All the landmark polygons which are expressed in the frame $\mathcal{F}_\mathcal{R}$ attached to O_1 and O_2. (c) All the landmark polygons which are expressed in the frame $\mathcal{F}_\mathcal{R}$ attached to O_2 and O_3.

putation is intensive.

In practice, two pairs of observations can be obtained from three observations in one straight trajectory. Figure 7a shows an example that three observations O_1, O_2 and O_3 are made along a straight trajectory with respect to three indistinguishable landmarks. The distance between O_1, O_2 and the distance between O_2, O_3 are not required to be equal. In this situation, three pairs of observations are obtained: O_1O_2, O_2O_3 and O_1O_3. Our method employs O_1O_2 and O_2O_3 to map the environment, O_1O_3 can be used for refining the map. The polygons derived from O_1O_2 and O_2O_3 are shown in Figs. 7b and 7c respectively. Since O_2 is a common observation in both O_1O_2 and O_2O_3, that is, data are associated with respect to O_2. Hence the complexity becomes $O((n!)^2)$.

4 Experimental Results

Our method was evaluated using a car-like robot (a lawn-mower tractor) in an outdoor environment. The robot was able to generate a video stream of the environment using its on-board omnidirectional camera. Three plastic cones with orange color were used as artificial landmarks L_a, L_b and L_c (see Fig. 8 of the experimental setup). There were other two natural landmarks L_d and L_e with orange color too. In the omnidirectional images, L_d and L_e were detectable due to their color. However, all the landmarks (including the orange cones) were visually indistinguishable in the images. Landmark bearings were derived from the omnidirectional images using color thresholding. The positions of all landmarks were measured manually and the robot's locations were estimated by an accurate localizer. The information of landmark positions and robot's trajectory were used for evaluating the accuracy of our SLAM method only, those information were unknown to the robot.

The robot was guided to move straight by visually locking a fixed target. In Fig. 8, the dotted track was the robot's moving trajectory. Three observation points O_1, O_2 and O_3 were randomly selected along the trajectory. Figure 9 shows the panoramic image captured at O_1.

At O_1, O_2 and O_3, the camera was able to detect L_c, L_d and L_e only, L_a and L_b were too far to detect. In Fig. 8, the
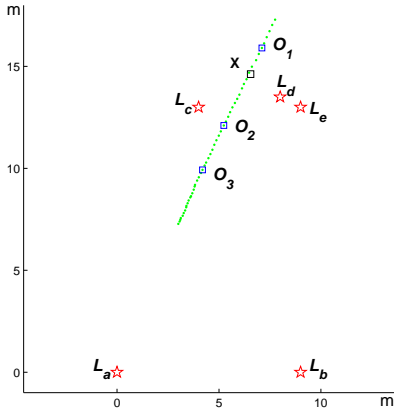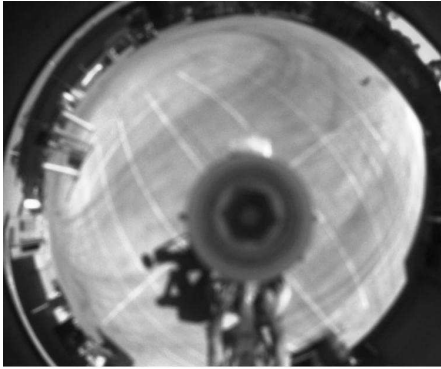
Figure 8: The experimental setup.

Figure 9: An example of the panoramic image captured at O_1.

Figure 10: The mapping results (with $\epsilon = 3$) in the $\mathcal{F}_\mathcal{L}$ attached to L_c and L_e.

visual order at O_1 was (L_d, L_e, L_c). When the robot moved from O_1 to the point X, the apparent angle with respect to L_d and L_e reduced gradually to 0 because X was collinear with L_d and L_e. From X to O_2, the indices in the visual order with respect to L_d and L_e were swapped. The visual order at O_2 became (L_e, L_d, L_c). By such a simple tracking, the landmark correspondences at O_1 and O_2 were found. Hence, the first bearing measured at O_1 and the second bearing measured at O_2 induced L_d. Similarly, the second bearing measured at O_1 and the first bearing measured at O_2 induced L_e. When the landmark correspondences are trackable, then two observations are sufficient for solving the SLAM problem.

When the landmark correspondences were not trackable due to the low frame rate, we employed the landmark bearings measured at O_1, O_2 and O_3 using the method described in Section 3 for the mapping. We varied the vision error ϵ from 1 to 7 degrees to analyse the sensitivity of the estimation accuracy of object positions. A sample output, using $\epsilon = 3$, in the static frame $\mathcal{F}_\mathcal{L}$ attached to L_c and L_e is presented in Fig. 10. The polygons represent the uncertainties of the ob-
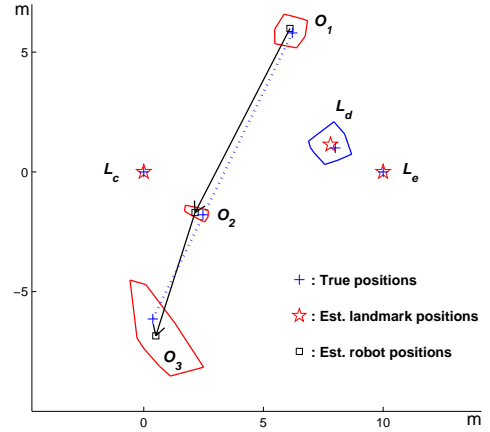
jects' positions. In these experiments the unit measurement in $\mathcal{F}_\mathcal{L}$ (the distance between L_c and L_e) is 10 m. Top chart and bottom chart of Fig. 11 present the experimental results of the location uncertainty and the estimation error respectively.

5 Conclusion

We have presented a bearing-only SLAM method for a 2D environment with indistinguishable landmarks. When the frame rate of the camera is high, the apparent angles and the visual orders derived from bearings are used for tracking the landmark correspondences. When the landmark correspondence is not trackable due to the low frame rate, we compute the most likely landmark positions from all possible hypotheses derived from the observations. The computation requires only landmark bearings derived from three observations along a straight trajectory.

The method proposed in this paper does not rely on object recognition techniques for solving the data association problem. In fact, most of the vision-based techniques for object recognition (such as SIFT [Lowe, 1999]) will not perform well if the landmarks are distant from the view point or their appearances are visually indistinguishable. Our SLAM method is useful in the case where the landmarks have exactly same appearance. The main advantage of our method is that it does not require odometry (the motion model) to predict robot's positions. Our method relies solely on bearing measurements (the observation model) only. The trade-off is that the robot has to move straight for a short while to initialize the landmarks. In practice, the robot's movement may not be perfectly straight. However, the non-straight nature can be compensated by increasing the confidence interval of the bearing measurements.

The complexity of our SLAM method with respect to n landmarks is $O((n!)^2)$. The computation of most likely landmark positions is straightforward, however, when n grows, the computation becomes expensive. At the initial stage of the SLAM process, a few visually distinctive landmarks are
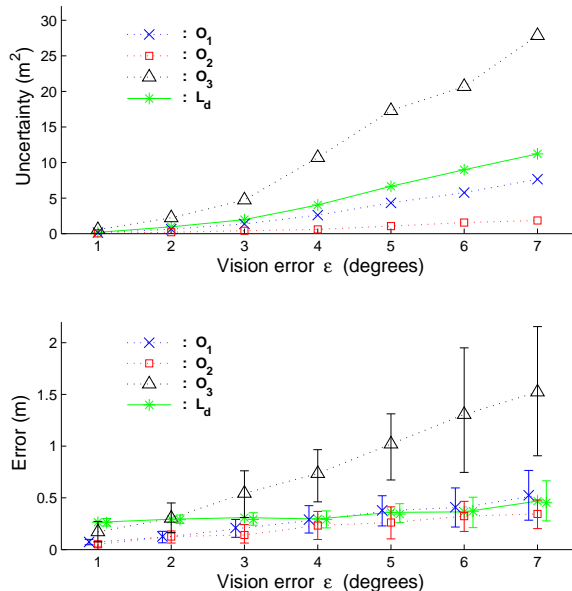
Figure 11: Top chart indicates that the locational uncertainties of O_1, O_2 and L_d are small. The uncertainty of O_3 is large because O_3 is distant from the landmarks and the movement between O_2 and O_3 is short. Bottom chart shows the error on the estimation of O_1, O_2, O_3 and L_d versus ϵ. The confidence intervals were derived from 40 experiments for a 80% confidence level.

sufficient for the mapping. Once an initial map with a few landmarks and robot's trajectory is built, new landmarks observed at subsequent observations can be added to the map incrementally. Since our mapping method requires no odometry nor range information, the induced map is up to a scale factor only. In practice, the scale factor can be obtained either from a cheap odometer or any range measurement in the environment, such as the distance between two landmarks.

In the future work, we will extend this work to deal with the situation where some of the landmarks are occluded.

Acknowledgments

References

[Bailey, 2003] T. Bailey. constrained initialization for bearing-only SLAM. *IEEE International Conference on Robotics and Automation*, 2003.

[Costa *et al.*, 2004] A. Costa, G. Kantor, and H. Choset. Bearing-only Landmark Initialization with Unknown Data Association. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, volume 2, pages 1764 – 1770, 2004.

[Davison *et al.*, 2004] A. Davison, Y. Cid, and N. Kita. Real-time 3D SLAM with wide-angle vision. *IAV2004 - 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, Lisboa Portugal*, 2004.

[Davison, 2003] A. Davison. Real-time simultaneous localization and mapping with a single camera. In *Proceedings of International Conference on Computer Vision, Nice, October*, 2003.

[Goncavles *et al.*, April 2005] L. Goncavles, E. di Bernardo, D. Benson, M. Svedman, N. Karlsson J. Ostrovski, and P. Pirjanian. A visual front-end for simultaneous localization and mapping. In *Proceedings of IEEE International Conference on Robotics and Automation, ICRA 2005*, pages 44–49, April 2005.

[Huang *et al.*, 2005] H. Huang, F. Maire, and N. Keeratipranon. Uncertainty Analysis of a Landmark Initialization Method for Simultaneous Localization and Mapping. In *Proceedings of Australian Conference on Robotics and Automation, Sydney*, 2005.

[Huang *et al.*, 2007] H. Huang, F. Maire, and N. Keeratipranon. Bearing-only Simultaneous Localization and Mapping for Vision-Based Mobile Robots. In Obinata G. and Dutta A., editors, *Vision Systems-Applications*, pages 335–360, Chapter 19. I-Tech Education and Publishing, Vienna, Austria, 2007. ISBN: 978-3-902613-01-1, available at: http://s.i-techonline.com/Book/Vision-Systems-Applications/ ISBN978-3-902613-01-1vsa19.pdf.

[Jensfelt *et al.*, 2006] P. Jensfelt, D. Kragic, J. Folkesson, and M. Bjorkman. A Framework for Vision Based Bearing Only 3D SLAM. *Centre for Autonomous System, Royal Institute of Technology, SE-100 44 Stockholm, Sweden*, 2006.

[Leonard and Durrant-Whyte, 1991] J. J. Leonard and H. F. Durrant-Whyte. Simultaneous localization for an autonomous mobile robot. In *Proceedings of IEEE/RSJ on Intelligent Robots and System, Osaka, Japan*, 1991.

[Lowe, 1999] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conferences on Computer Vision, Corfu, Greece*, 1999.

[Spero, 2005] D. J. Spero. *Simultaneous Localisation and Map Building: The Kidnapped Way*. PhD thesis, Intelligent Robotics Research Centre, Monash University, Melbourne, Australia, 2005.

[Zunino and Christensen, 2001] G. Zunino and H. I. Christensen. Simultaneous localization and mapping in domestic environments. In *Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent System*, 2001.