Bearing-only Simultaneous Localization and Mapping for Vision-Based Mobile Robots

Henry Huang, Frederic Maire and Narongdech Keeratipranon
Faculty of Information Technology, Queensland University of Technology
Australia

1. Introduction

To navigate successfully, a mobile robot must be able to estimate the spatial relationships of the objects of interest accurately. A SLAM (Simultaneous Localization and Mapping) system employs its sensor data to build incrementally a map of an unknown environment and to localize itself in the map simultaneously. Thanks to recent advances in computer vision and cheaper cameras, vision sensors have become popular to solve SLAM problems (Bailey, 2003; Costa et al., 2004; Davison et al., 2004; Goncavles et al., 2005; Jensfelt et al., 2006; Mouragnon et al., 2006). The proposed bearing-only SLAM system requires only a single camera which is simple and affordable for the navigation of domestic robots such as autonomous lawn-mowers and vacuum cleaners.

Solutions to SLAM problems when the mobile robot is equipped with a sensor that provides both range and bearing measurements to landmarks are well developed (Leonard & Durrant-Whyte, 1991; Zunino & Christensen, 2001; Spero, 2005; Bailey & Durrant-Whyte, 2006). With a single camera, landmark bearings can be derived relatively easily from a grabbed image, however it is much more difficult to obtain accurate range estimates. Due to the low confidence in range estimates from vision data, it is desirable to solve SLAM problems with bearing only measurements.

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. A typical bearing-only SLAM system requires multiple observations for landmark initialization through triangulation. With only one observation, a stereo vision can provide range measurements because its multiple cameras grab images from slightly different viewpoints. However the reliable vision range in a stereo vision is limited due to the distance between the two cameras. Several observations at different locations are required to provide a robust range estimate.

Structure From Motion (SFM) is a process to construct the map of an environment with the video input from a moving camera. SFM allows a single camera to grab images at some vantage points for landmark initialization, such as a sufficient baseline and a straight movement. The requirement of SFM is well satisfied with a mobile robot, some recent works had utilized SFM to bearing-only SLAM (Goncavles et al., 2005; Jensfelt et al., 2006). Our method to bearing-only SLAM is inspired from the techniques used in both stereo vision and SFM.
Existing approaches to bearing-only SLAM require the readings from an odometer to estimate the robot locations prior to landmark initialization. It can be argued that such approaches are not strictly bearing-only SLAM as they rely on odometric information. This chapter presents a new 2-dimensional bearing-only SLAM system that relies only on the bearing measurements from a single camera. Our proposed system does not require any other sensors like range sensors or wheel encoders. The trade-off is that it requires the robot to be able to move in a straight line for a short while to initialize the landmarks. Once the landmark positions are estimated, localization becomes easy. The induced map created by our method is only determined up to a scale factor as only bearing information is used (no range or odometry information). All the object coordinates in the map multiplied by a scale factor would not change the bearing values.

The structure of this chapter is as follows. First, we introduce a direct localization method using only the bearings extracted from two panoramic views along a linear trajectory. We explain how to induce a Cartesian coordinate system with only two distinguishable landmarks. The method is then extended to landmark initialization with more landmarks in the environment.

In general, vision sensors are noisy. Dealing with sensory noise is essential. Two different methods are presented to compute the spatial uncertainty of the objects:

1. A geometric method which computes the uncertainty region of each landmark as the intersection of two vision cones rooted at the observation points.
2. A probabilistic method which computes the PDFs (Probability Density Functions) of the landmark positions. Formulas are derived for computing the PDFs when an initial observation is made.

The proposed SLAM system requires only a single camera, an interesting setup for domestic robots due to its low cost. It can be fitted to a wheeled robot as well as a legged robot.

2. Related work

The term SLAM was first introduced by Leonard and Durrant-Whyte (1991), it refers to Simultaneous Localization and Mapping. SLAM is one of the fundamental tasks in the navigation of an autonomous mobile robot. In robotic navigation, a map is a representation of the spatial relationship between the objects of interest in the environment. A map usually contains the positions of certain objects of interest, such as landmarks and obstacles. The process of a robot to determine its position in a map is called localization. GPS (Global Positioning System) is a popular localization system, in which the map is given for navigation. GPS is well suited for vehicles to navigate in a large scale outdoors environment, for instance, to navigate from city to city. For a domestic robot, however, a GPS is not accurate enough and does not work properly indoors and in some built-up areas. Further, the map of a particular environment may not be always available. A domestic robot cannot localize itself without a map. A SLAM system needs to build incrementally a map while it explores the environment and to determine its location in the map simultaneously.

For localization the robot needs to know where the landmarks are, whereas to estimate the landmark positions the robot needs to know its own position with respect to the map. The problem of SLAM is considered as a “chicken-and-egg” problem (Siegwart & Nourbaksh, 2004). To predict the position of the robot, conventional SLAM systems rely on odometry. Unfortunately, the accumulation of odometric errors (due mainly to wheel slippage) makes
the accuracy of the position estimates based only on odometry deteriorate rapidly. Updating the estimates with other sensory input is needed if the robot navigates for a long time. Solutions to SLAM can be found if both range and bearing measurements to landmarks are obtained through sensors (Leonard & Durrant-Whyte, 1991; Zunino & Christensen, 2001; Spero, 2005; Bailey & Durrant-Whyte, 2006). Such a sensor reading refers to a Full Observation. A full observation can be achieved by either a single sensor (i.e., a laser range finder) or a combination of two sensors (i.e., a sonar sensor and a camera). Range and bearing measurements constitute a full state of the environment. The sensors which observe the full state of the environment (i.e., both range and bearing) are called range-bearing sensors. A full observation is sufficient to form an estimate, such as an uncertainty region, of a landmark position. A typical uncertainty region is a Gaussian distribution over the possible positions of a landmark. Updating an estimate can be achieved by fusing the estimates from the subsequent observations. However, a range-bearing sensor is too expensive for a domestic robot. Solving the SLAM problems with a cheaper sensor is desirable. A sensor reading with either range-only or bearing-only measurement to a landmark is called a Partial Observation. A partial observation is insufficient to completely determine a landmark position. A partial observation generates only a non-Gaussian distribution over an unbounded region for the landmark position (Bailey & Durrant-Whyte, 2006). Multiple observations from several vantage points are required to estimate the landmark position. A sensor reading obtained from a single camera constitutes only a partial observation because it provides bearing measurements but does not provide accurate range measurements. In general, a vision sensor is relatively cheaper than a range-bearing sensor. We wish to solve SLAM problems with bearing-only measurements. Next section reviews related work on vision based navigation for bearing-only SLAM.

2.1 Vision based navigation
Vision based navigation for a mobile robot had been investigated since early nineties of last century. Levitt and Lawton (1990) developed a Qualitative Navigator based on vision sensors. This navigator was able to explore the environment and to determine the relative positions of all the objects of interest. In general, an image contains very rich information for mapping the corresponding environment. A certain feature can be recognized through its specific color, shape and size. The frame rate up to 30 Hz from a video camera also enhances to SLAM, in particular to solve the data association problem. Landmark bearings can be derived from a panoramic image taken by an omni-directional vision sensor (for example, a single camera aiming at a catadioptric mirror). A panoramic image offers a $360^\circ$ view of the environment. Because of the robustness of bearing estimates and the complete view of the environment, previous works have utilized omni-directional vision sensors in robotic navigation (Rizzi & Cassinis, 2001; Usher et al., 2003; Menegatti et al., 2004; Huang et al., 2005b). Stereo vision is another option used in robotic navigation. In addition to the bearing information, a stereo vision sensor can also measure the depth to a landmark (Murray & Jennings, 1998; Se et al., 2002; Sabe et al., 2004). A typical stereo vision sensor consists of two cameras, also known as a Binocular Vision. The disparity of the images taken from two slightly different viewpoints determines the landmark's range through triangulation. A Baseline in stereo vision is a line segment connecting the centres of two cameras' lens. Some stereo vision systems consist of three cameras, they are called Trinocular Visions.
configuration of a trinocular vision is to put three cameras on a right angle polygonal line. A trinocular vision can achieve better results than a binocular vision because the second pair of cameras can resolve situations that are ambiguous to the first pair of cameras (Se et al., 2002; Wooden, 2006).

The length of the baseline is essential in stereo vision, because it affects the precision of depth estimation and the exterior design of robotic hardware. QRIO (Sabe et al., 2004), a humanoid robot having a 5cm baseline in its binocular vision. The error of the depth measurement at a distance of 1.5m is over 80mm. The depth estimates of objects with the distances of 2m or more are omitted. LAGR (Wooden, 2006), an outdoor robot equipped with a trinocular vision of Point Grey Bumblebee. A maximum vision range of 6m was reported with a baseline of 12cm. To maximize the vision range of a stereo vision sensor, a longer baseline is required. Based on the mobility of a mobile robot, it is possible to extend the distance of any two viewpoints of a single camera (called a Monocular Vision). If a robot can move straight, the estimation of a landmark range from a monocular vision will be the same as the estimation in a stereo vision. Such approach was first proposed by Huang et al. (2005a). In this paper, a localization method with two observed bearings along a linear trajectory was presented. The method is particularly useful and accurate if the robot can move straight, i.e., the robot’s yaw is toward to a specific landmark.

In computer vision, Structure From Motion (SFM) refers to the process of building a 3D map of a static environment from the video input from a moving camera. This is very similar to stereo vision where a 3D map is built from two simultaneous images of the same landmark. In both cases, the same landmark is taken into multiple images and the disparities of images are used to compute the landmark location. In stereo vision, the images are taken at different viewpoints simultaneously. In SFM, due to the robot’s motion, the images are taken at different viewpoints at different time steps. Visual odometry (Nister et al., 2004) employs SFM to estimate the motion of a stereo head or a single moving camera based on video data. The front end of this system is a feature tracker. Point features are matched between pairs of frames and linked into image trajectories at video rate. SFM presents significant advantages compared with a stereo vision due to the low cost of a monocular vision and the flexible baseline. However, SFM can build a map with respect to a static environment only because of the images are obtained at different time steps. Goncavles et al. (2005) presented a framework to bearing-only SLAM based on SFM from three observations. They utilized a wall corner as the landmark for guiding the robot to move straight. Three images were taken while the robot was moving toward the wall corner. Each image was taken when the robot had moved 20cm approximately. A similar work (Jensfelt et al., 2006) was using \( N \) images for landmark initialization, here \( N \) is a sufficient number to obtain an accurate estimation. To ensure a proper triangulation, the images were discarded if the robot had not moved more than 3cm (i.e., baseline under 3cm) or turned more than 1 degree (i.e., not a straight movement). Both of the approaches solve the bearing-only SLAM problem using a monocular vision. However, they require an odometer to determine robot’s motion. Our method to bearing-only SLAM is similar to SFM with a monocular vision, but does not rely on odometric information.
2.2 Dealing with uncertainty

In general, vision sensors are noisy. Dealing with sensory noise is essential in robotic navigation. The uncertainty of an object location can be represented with a PDF (Probability Density Function). When a robot is initially placed in an unknown environment without any prior information, the PDF of the robot position is uniform over the whole environment. Once the robot starts to sense the environment, information gathered through the sensors is used to update the PDF. Smith and Cheeseman (1986) estimated the object locations by linking a series of observations through an approximate transformation. The transformation includes compounding and merging the uncertain spatial relationships from sensor readings. Stroupe et al. (2000) showed how to fuse a sequence of PDFs of 2-dimensional Gaussians estimated from noisy sensor readings.

Robustness to sensory noise can be achieved with probabilistic methods such as Extended Kalman Filters (EKF) or Particle Filters (PF). The PF follows the Sampling Importance Resampling (SIR) algorithm, also known as the Monte Carlo Localization (MCL) algorithm in robotics (Fox et al., 1999). In PF, the number of particles is an important factor to the computing complexity. Montemerlo et al. (2003) proposed an efficient algorithm called FastSLAM based on PF with a minimized number of particles. Davison (2003) used a separate PF to estimate the distance from the observation point to the landmark with a single camera. The estimated distance is not correlated with other observations due to the limitation of the field of view. The follow-up work (Davison et al., 2004) improved the SLAM results by applying a wide-angle camera. In (Menegatti et al., 2004), omnidirectional images were employed to the image-based localization combining with MCL technique. Sim et al. (2005) solved SLAM problem with PF using a stereo vision sensor.

EKF is computationally efficient for positional tracking. However, an initial estimate of Gaussian distribution over the landmark position is required. This estimate can be refined efficiently with the estimates from subsequent observations. It is important to have an initial estimate relatively close to the real solution. Many works have focused on the problem of landmark initialization. In (Bailey, 2003), previous poses of the robot were stacked in the memory to perform the landmark initialization. Once the landmarks were initialized, the batch of observations was used to refine the whole map. Costa et al. (2004) presented an iterative solution to the landmark initialization of bearing-only SLAM problem with unknown data association (i.e., all landmarks are visually identical). The authors estimated landmark positions through Gaussian PDFs that were refined as new observations arrived.

Bundle adjustment is a process which adjusts iteratively the robot’s pose and the landmark positions in order to obtain the optimal least squares solution. Combining EKF with bundle adjustment ensures a robust estimate. Such an optimization is usually calculated off line due to expensive in computation. In (Mouragnon et al., 2006), landmark initialization was carried out with a bundle adjustment in an incremental way, in the order of video frames. An incremental method can improve the computing efficiency compared with the usual hierarchical method.

Landmark initialization based on memorizing previous measurements or iterative methods cause time delay for estimation. These methods belong to the delayed methods of landmark initialization (Sola et al., 2005). Some immediate initialization methods to bearing-only SLAM called undelayed methods of landmark initialization were introduced; Kwok and Dissanayake (2004) presented a multiple hypotheses approach to solve the problem in a computationally efficient manner. Each landmark was initialized in the form of multiple
hypotheses distributed along the direction of the bearing measurement. The validity of each hypothesis was then evaluated based on the Sequential Probability Ratio Test (SPRT). Sola et al. (2005) gave a new insight to the problem and presented a method by initializing the whole vision cone (see Figure 4(a)) that characterizes the direction of the landmark. This cone is covered by a sequence of ellipses that represent the likelihood of the landmark. Undelayed method of landmark initialization is efficient to identify the directions of all landmarks when the first bearing measurements are made. It does not state explicitly the locations of the landmarks. Further observations are still required to initialize the landmark positions. Lemaire et al. (2005) applied an undelayed initialization method to a 3D bearing-only SLAM. The landmark initialization is similar to the method proposed in (Kwok & Dissanayake, 2004) by maintaining a mixture of Gaussians. The updating process was done by comparing the likelihoods of subsequent observations. If the likelihood falls below a certain threshold then the Gaussian is removed. Once only a single Gaussian is left in the cone, the landmark is initialized and added into the map for EKF-SLAM.

2.3 Our approaches
This chapter presents two methods to compute the spatial uncertainties of the objects based solely on bearing measurements only: namely a geometric method and a probabilistic method. These methods are similar to the approach of the undelayed method of landmark initialization. Since the estimate based on a partial observation (known bearing but unknown range) is insufficient to completely determine a landmark position, a second observation from a vantage position is required to generate an explicit estimate.

In the geometric method, we manipulate directly each vision cone as a polyhedron instead of a sequence of Gaussians. Each cone contains a landmark position. After a second observation in a linear trajectory, the uncertainty region (the set of possible landmark positions that are consistent with the first and second observations) of the landmark becomes the intersection of two cones rooted at the two observation points, see Figure 4(b). Depending on the difference of bearings, the intersection is either a quadrangle (four-side polygon) or an unbounded polyhedron. For each estimation, we change the bases from the local frame (the robot-based frame, denoted by \( R_F \)) into the global frame (the landmark-based frame, denoted by \( L_F \)). The uncertainties of all objects are re-computed with respect to \( F_i \) by the change of bases. A global map with the estimated positions of all objects and their associated uncertainties can be gradually built while the robot explores its environment.

In the probabilistic method, a landmark position is represented by a PDF \( p(r, \alpha) \) in a polar coordinate where \( r \) and \( \alpha \) are independent. Formulas are derived for computing the PDF of landmark position when an initial observation is made. The updating of the PDF with the subsequent observations can be done by direct computing from the formulas. We select a number of sample points in the uncertainty region (computed from the geometric method) by the rejection method (Leydold, 1998). These sample points are used to represent the PDF in \( F_i \). By changing the bases from \( F_i \) to \( F_j \), the PDFs of all object positions in the global frame \( F_j \) can also be computed.

Without range measurement, we assume the probability density of an object location is constant along the range. It is a more realistic assumption than the one made by other existing methods (Davison, 2003; Davison et al., 2004; Kwok & Dissanayake, 2004; Sola et al., 2005) which assume that the probability density of the object location is a Gaussian or a
mixture of Gaussians. Indeed, if only bearing information is given, the probability that the landmark is between 4 and 5 metres should be the same as the probability that the landmark is between 5 and 6 metres. The representation with a Gaussian or a mixture of Gaussians fails in this constraint. With our PDF representation, the probability that the landmark is between 4 and 5 metres will be the same as the probability that the landmark is between 5 and 6 metres.

3. A direct localization method using only the bearings extracted from two panoramic views along a linear trajectory

In this section, we describe a direct method (in the sense it does not use an iterative search) based solely on vision for localizing a mobile robot in a 2-dimensional space. This method relies only on the bearings derived from two images taken along a linear trajectory. We only assume that the robot can visually identify landmarks and measure their bearings. This method does not require any other sensors (like range sensors or wheel encoders) or the prior knowledge of relative distances among the objects. This method can be adopted in a localization system which utilizes only a single camera as the sensor for navigation. Given its low cost, such a system is well suited for domestic robots such as autonomous lawn-mowers and vacuum cleaners.

3.1 Method description

In order to describe our method we need to introduce some notation. The robot position at \( i \)th observation point is denoted by \( O_i \). The position of \( j \)th landmark is denoted by \( L_j \). The notation \( \beta \) denotes the bearing measurement at \( O_i \) with respect to \( L_j \). The line going through two points \( x_1 \) and \( x_2 \) will be denoted by \( \Gamma(x_1,x_2) \). This section shows how to compute the Cartesian coordinates of landmarks and the robot positions from the bearings measured at \( O_1 \) and \( O_2 \) relatively to \( L_1 \) and \( L_2 \). We consider two right-handed coordinate systems, the robot-based frame denoted by \( R_F \), and the landmark-based frame denoted by \( L_F \). In Figure 1(a), the coordinates of \( O_1 \) and \( O_2 \) in \( R_F \) are respectively \([0,0]^T\) and \([0,1]^T\). Similarly, in \( L_F \) (see Figure 1(b)), the coordinates of \( L_1 \) and \( L_2 \) are respectively \([0,0]^T\) and \([1,0]^T\). The frame \( F_L \) is a global frame since all the landmarks are assumed static in the environment. The distance \( \|L_1 - L_2\| \) is taken as the measurement unit for the localization system.

While the robot moves in a linear trajectory, two images are taken at \( O_1 \) and \( O_2 \) respectively. The landmark bearings are derived from these two images. The position \( L_1 \) in \( F_x \) is computed as the intersection of two lines \( \Gamma(O_1,L_1) \) and \( \Gamma(O_2,L_1) \). The equations of this two lines can be obtained from the bearings \( \beta_1 \) and \( \beta_2 \), and the coordinates of \( O_1 \) and \( O_2 \) in \( F_x \). Once \( L_1 \) is available in \( F_x \), we can determine the affine transformation that relates the coordinates \( X_1(x) \) and \( X_2(x) \) of a point \( x \) in the two coordinate systems \( F_1 \) and \( F_x \). That is, an expression of the form \( X_1(x) = a^* X_2(x) + b \), here \( a \) is a matrix and \( b \) is a vector. The coordinates of \( O_1 \) and \( O_2 \) in \( F_x \) are then easily derived.
Figure 1. From two landmark bearings observed at points $O_1$ and $O_2$, the coordinates of $L_1$ and $L_2$ in $F_R$ are computed. Then a simple change of bases gives the positions of $O_1$ and $O_2$ in the global frame $F_L$

Figure 2. The estimated error $e$ on $L_2$ depends on the relative difference in bearings

In order to determine the relative position of a landmark, this landmark should not be on the line $\Gamma(O_1, O_2)$. For example, if $L_1$, $O_1$, and $O_2$ are on the same line, then $\Gamma(O_1, L_1) \cap \Gamma(O_2, L_1)$ is not a single point but a whole line.

Experiments in simulation and on a real robot (see Section 3.2) indicate that the accuracy of this localization system is sensitive to the relative difference of bearings. Let $e$ be the
estimated error of landmark position. Figure 2 shows that \( d_e = d \sin(\beta^2) = e \sin(\theta) \), here \( \theta = \beta^2 - \beta^1 \), \( d_o \) is the distance between \( L_i \) and \( O \), and \( d = \| O_i - O_o \| \). We have \( e = \frac{\sin(\beta^2)}{\sin(\theta)} \). That is, \( e \) is proportional to the landmark range \( d \) and the inverse of the relative difference in bearings. Assume the landmark range is fixed and the bearings angles are small, the ratio \( \frac{e}{d} \) will be approximately equal to \( \frac{\beta^2}{\beta^2 - \beta^1} \). This result confirms our intuition that a large relative change in bearings should give a more accurate position estimate.

3.2 Empirical Evaluation

Our localization method was evaluated on a Khepera robot equipped with a color camera (176 x 255 resolution). The average error between the measured and actual bearings is about \( \pm 2 \) degrees. In this experiment, the second landmark was placed 20 centimetres apart from the first landmark. Four different starting points were used, and 20 trials at each point were conducted. The moving distance in all cases was 30 centimetres. The moving directions were westwards parallel to the landmarks. The results are shown in Figure 3. In this figure, landmarks are denoted by stars, trajectories are shown as arrows, and the estimated positions by our localization method are displayed as scatter points.

The localization error, average distances between the estimated positions and the actual positions, at positions \( a, b, c, \) and \( d \) (in Figure 3) are respectively 0.6, 1.2, 2.2, and 2.8 centimetres. The errors are small compared to the diameter of the robot (6 centimetres). Other experimental results have confirmed that the error is inversely proportional to the relative difference in bearings.

![Figure 3. Estimated positions of the robot determined by the proposed localization method](image)

When more than two landmarks are present, the localization accuracy can be further improved by fusing the estimated positions, giving more importance to the estimation returned by the pair of landmarks that has a larger relative difference in landmark bearings. In summary, we have introduced a novel effective approach for robot self-localization using only the bearings of two landmarks. This technique can be viewed as a form of stereo-vision. The method we propose is well suited for real-time system as it requires very little computation.
When more than two landmarks exist in the environment, the robot can determine the relative positions of the landmarks provided some weak visibility constraints are satisfied. Indeed, suppose there are two pairs of landmarks \( \{L_1, L_2\} \) and \( \{L_3, L_4\} \) visible from a segment \( O_1O_2 \) (notice that \( \{L_1, L_2\} \) and \( \{L_3, L_4\} \) do not have to be in direct line of sight). Then using three different bases, the first one \( B_{1} \) attached to \( O_1O_2 \), the second one \( B_{2} \) attached to \( \{L_1, L_2\} \), and the third one \( B_{3} \) attached to \( \{L_3, L_4\} \), we can determine the change of basis matrices \( M_{B_{2}B_{1}} \) and \( M_{B_{3}B_{1}} \). The matrix product \( M_{B_{3}B_{1}}M_{B_{2}B_{1}}^{-1} \) allows us to compute the positions of the pairs of landmarks \( \{L_1, L_2\} \) and \( \{L_3, L_4\} \) relatively to each other.

This method enables a mobile robot to localize itself with only two observed bearings of two landmarks. Such a localization system will be invaluable to an indoor robot as well. As the bearings of the sides of a door frame can play the roles of the landmarks \( L_1 \) and \( L_2 \) and tell the robot exactly where it stands relative to the door. In next section, we employ this method to solve the landmark initialization problem in bearing-only SLAM.

4. Sensitivity Analysis to Landmark Initialization of Bearing-Only SLAM
– A geometric approach

In this section, we propose a geometric method to solve the landmark initialization problem in bearing-only SLAM. The assumptions and the localization method are the same as in Section 3, with the exception that vision error is taken into consideration. The estimate of a landmark position becomes an uncertainty region instead of a single point. In particular, we show how the uncertainties of the measurements are affected by a change of frames. That is, we determine what can an observer attached to a landmark-based frame \( F_L \) deduced from the information transmitted by an observer attached to the robot-based frame \( F_R \).

4.1 Method description

The notations in this section are the same as in Section 3.1. The uncertainty region of \( L_j \) is denoted by \( A_j \). Assume that the error range for the bearing is \( \pm \epsilon \). In other words, at an observation point \( O_i \), a landmark position \( L_j \) is contained in the vision cone which is formed by two rays rooted at \( O_i \). The first ray is defined by \( O_i \) and the bearing \( \beta_i + \epsilon \); the second ray is defined by \( O_i \) and bearing \( \beta_i - \epsilon \). Figure 4(a) shows the vision cones in the robot-based frame \( F_R \) based on the reading of the landmark bearings from \( O_i \). After reading the bearing measurements from both \( O_i \) and \( O_j \), the uncertainty region \( A_i \) becomes the intersection of two cones rooted at \( O_i \) and \( O_j \) respectively. Figure 4(b) shows that a typical intersection is a 4-sided polygon. If the cones are almost parallel, their intersection can be an unbounded polyhedron.

The spatial relationships in Figure 4(b) are expressed in \( F_R \). Since the robot is moving over time, the base of \( F_R \) changes too. Therefore, it is necessary to change coordinate systems to express all positions in the global frame \( F_L \). Figure 5 illustrates the difficulty of expressing
the robot centred information in the global frame $F_x$. The uncertainty on the landmarks prevents us from applying directly a change of bases. In next section, we will show how to solve this problem.

Figure 4. (a) The vision cones rooted at $O_1$ contain the landmarks. Each cone represents the unbounded uncertainty of the associated landmark. The diagram is drawn with respect to the robot-based frame $F_R$. (b) The intersections of the vision cones form the uncertainty region $A_i$.

Figure 5. When $L_1$ and $L_2$ are not certain, a simple change of bases does not induce correct uncertainty regions of $L_3$ and $L_4$. 
4.2 Uncertainty and change of frames

From the uncertainties of landmark positions estimated in the robot-based frame \( F_R \), we would like to derive the uncertainty regions of the observed objects with respect to the landmark-based frame \( F_L \). Given a point \( x \), if \( X_R(x) \) and \( X_L(x) \) denote the coordinate vector of \( x \) in frames \( F_R \) and \( F_L \) respectively.

Consider the simple case of Figure 6(a) which contains only two landmarks and two robot positions. Assume the robot (the observer) sees \( L_1 \) clearly from \( O_1 \) and \( O_2 \), but sees \( L_2 \) with some noise. The uncertainty region of \( L_1 \) in \( F_R \) is reduced to a single point (no ambiguity). Whereas, the uncertainty region of \( L_2 \) in \( F_R \) is a polyhedron.

![Figure 6. (a) A simple case in \( F_R \), where we assume \( L_1 \) is clearly observed without ambiguity, the uncertainty region of \( L_2 \) is \( A_{L_2} \). The four vertices of \( A_{L_2} \) are denoted by \( L_1^{k} \), \( k = 1...4 \). (b) After the change of frames, the uncertainty regions of \( O_1 \) and \( O_2 \) are denoted by \( A_{O_1} \) and \( A_{O_2} \). We obtain \( O_k^R \) from Equation (4) with respect to \( L_2^{k} \), \( k = 1...4 \). The uncertainty regions of \( O_1 \) and \( O_2 \) with respect to \( F_L \) can be obtained by considering all possible hypotheses for the location of \( L_2 \), consistent with the observations. That is, we consider the set of possible coordinate vectors \( X_R(L_2) \) of \( L_2 \) in \( F_R \). For each hypothesis \( X_R(L_2) = h_2 \), a standard change of bases returns the coordinates \( X_L(O_1) \) and \( X_L(O_2) \) of respectively \( O_1 \) and \( O_2 \) in \( F_L \). Making \( h_2 \) range over the vertices of \( A_{L_2} \) in Figure 6(a), create the polyhedra \( A_{O_1} \) and \( A_{O_2} \) of uncertainty regions with respect to \( F_L \) (see Figure 6(b)).
In the general case, when uncertainty exists for both $L_1$ and $L_2$, to transfer the information from $F_R$ to $F_L$, we consider simultaneously all the possible locations of $L_1$ and $L_2$ consistent with the observations. We hypothesize:

\[ X_R(L_1) = h_1, \quad \text{and} \quad X_R(L_2) = h_2 \]  

Let $\tau_{h_1, h_2}$ be the affine transformation function for changing frames from $F_R$ to $F_L$. That is,

\[ X_L(L_1) = \tau_{h_1, h_2}(X_R(L_1)) = [0, 0]^T \]  

\[ X_L(L_2) = \tau_{h_1, h_2}(X_R(L_2)) = [1, 0]^T \]

The above constraints completely characterize $\tau_{h_1, h_2}$. For any point $x$, the coordinates transfer between the two frames is done with Equation (4).

\[ X_L(x) = \tau_{h_1, h_2}(X_R(x)) \]  

In other words, the uncertainty region $A_{0_1}$ of the robot position in $F_L$ is

\[ X_L(A_{0_1}) = \bigcup_{h_1, h_2 \in A_1} \tau_{h_1, h_2}(X_R(O_i)) \]

Figure 7. The uncertainty regions in $F_L$ are derived from the uncertainty regions in $F_R$ (see Figure 5). The centroids of the uncertainty regions are used to represent the estimated positions of different objects. The areas of the polyhedra quantify how uncertain the estimates are.

The uncertainty regions $A_{0_1}$ for $L_1$ and $L_4$ in $F_L$ are computed similarly.
\[ X_i(A_i) = \bigcup_{k \in A_i} \tau_{k,i}(X_k(A_k)) \tag{6} \]

The computation in this example, we take the four vertices (the extreme points) \( L_1 \) and \( L_2 \) \((k = 1...4)\) from \( A_i \). Figure 7 shows the estimated uncertainties of \( O_1 \), \( O_2 \) and \( L_1, L_2 \) in \( F_i \). The polyhedron \( X_i(A_{i_{OL}}) \) approximates the set of all consistent points for \( O_i \) and the polyhedron \( X_i(A_{i_{AL}}) \) approximates the set of all consistent points for \( L_i \). Although the uncertainty region \( X_i(A_{i_{OL}}) \) is not a polyhedron, in practice it can be approximated by a polyhedron. We have tested the proposed method both in simulation and on a real robot. These results are presented in next section.

4.3 Simulation

We tested the proposed method in simulation in an environment with four landmarks (at unknown positions to the localization system). The robot moves in a polygonal line around the centre with some randomness. Since we focus on landmark initialization, Figure 8 shows only the estimated positions of the landmarks.

Figure 8. The uncertainty regions \( A_{i_{OL}} \) and \( A_{i_{AL}} \) gradually shrink as the number of observations increases. The arrows represent the robot movements.

Two landmarks are arbitrarily selected as \( L_i \) and \( L_j \). With the change of frames from \( F_i \) to \( F_j \), the uncertainty regions \( A_{i_{OL}} \) and \( A_{i_{AL}} \) are computed. When another pair of observations is
available after the robot has moved again, new $A_i$ and $A_k$ are obtained in the same manner. The estimated positions from all movements are unifiable since they are with respect to the same frame $F_i$. Figure 8 shows how the uncertainty regions are refined after several movements. The polyhedra $A_i$ and $A_k$ shrink gradually. A global map with the estimated positions and the corresponding uncertainties of all landmarks can be incrementally built.

4.4 Evaluation on a Real Robot

Our method was evaluated using a Khepera robot. The Khepera robot has a 6 centimetre diameter and is equipped with a color camera (176 x 255 resolution). A kheperaSot robot soccer playing field, 105 x 68 square centimetres, was used as the experimental arena (see Figure 15). There were four artificial landmarks in the playing field. Only one landmark was distinct from the others. The second landmark was placed 20 centimetres apart from the first landmark.

During the experiments, the robot moved in a polygonal line. Two panoramic images were taken in each straight motion. Landmark bearings were extracted from the panoramic images using a color thresholding technique. Bearings from each pair of observations were used to estimate the landmark positions. The vision error $\varepsilon$ is limited to $\pm2$ degrees.

Figure 9(a) shows the estimated uncertainties of landmark positions after 10 pairs of observations. The actual landmark positions are denoted by stars, the estimated landmark positions are shown as circles, and the areas of the polyhedra represent the uncertainties.

The uncertainties of $L_1$ and $L_4$ decrease rapidly in the first few observations and does not change much after the third observation as shown in the top chart of Figure 9(b). The bottom chart of Figure 9(b) displays the estimated errors of $L_2$ and $L_4$ are 2 centimetres and 3 centimetres respectively. The measurement unit in this chart equals to 20 centimetres.
We carried out another experiment to study the sensitivity of vision error $\varepsilon$. The top chart of Figure 10 shows the relationship between $\varepsilon$ and the uncertainties of $L_i$. The amount of the vision error $\varepsilon$ was varied from 2 to 7 degrees. The uncertainties are proportional to $\varepsilon$ in a linear manner. The bottom chart of Figure 10 shows that the estimated error might not decrease monotonically. This is because we assign the centroid of the uncertainty region as the estimated landmark position.

In this section, we introduced a method for analyzing how uncertainty propagates when information is transferred from one observer attached to a robot-based frame to an observer attached to a landmark-based frame. The accuracy of this method was demonstrated both in simulation and on a real robot. In next section we will employ a probabilistic method to compute the uncertainties of object positions.

![Figure 10. Uncertainties and estimated errors at different amounts of vision error $\varepsilon$ in degrees](image)

### 5. Sensitivity Analysis to Landmark Initialization of Bearing-Only SLAM – A probabilistic approach

In this Section, we describe a probabilistic method to solve the landmark initialization problem in bearing-only SLAM. The assumptions in this method are the same as Section 3. We characterize $p(r, \alpha)$ the PDF of landmark position expressed in polar coordinates when
5.1 Method description

Let \( p(r, \alpha) \) be the PDF of the landmark position in polar coordinates when only one observation has been made. We characterize \( p(r, \alpha) \) when \( r \) and \( \alpha \) are independent. Let \( \beta \) denote the measured landmark bearing. Assume that the error range for the bearing is \( \pm \varepsilon \). The landmark position is contained in the vision cone which is formed by two rays rooted at the observation point with respect to two bearings \( \beta - \varepsilon \) and \( \beta + \varepsilon \) (see Figure 11).

![Figure 11. The vision cone is rooted at the observation point. The surface of the hashed area is approximately \( r \, \text{d}r \, \text{d}\alpha \) for small \( \text{d} \alpha \) and \( \text{d}r \).](image)

The surface of the hashed area in Figure 11 for small \( \text{d}r \) and \( \text{d} \alpha \) can be computed as

\[
\left[ \pi (r + \text{d}r)^2 - \pi r^2 \right] \frac{\text{d} \alpha}{2\pi} = \frac{1}{2} \left[ 2r \, \text{d}r + (\text{d}r)^2 \right] \text{d} \alpha \equiv r \, \text{d}r \, \text{d} \alpha
\]

Because the probability of the landmark being in the vision cone is 1, we have

\[
\int_{\Delta} \int_{-\varepsilon}^{\varepsilon} p(r, \alpha) r \, \text{d}r \, \text{d} \alpha = 1
\]  

In Equation (7), \( R_{\text{min}} \) and \( R_{\text{max}} \) are the bounds of the vision range interval. We define \( F(R) \) as the probability of the landmark being in the area \( \{(r, \alpha) : r \in [R_{\text{min}}, R], \alpha \in [\beta - \varepsilon, \beta + \varepsilon]\} \), \( F(R) \) can be represented as:

\[
F(R) = \int_{\Delta} \int_{-\varepsilon}^{\varepsilon} p(r, \alpha) r \, \text{d}r \, \text{d} \alpha
\]
We define $\Psi(R, \Delta)$ as the probability of the landmark being in the dotted area in Figure 11. Since $\Psi(R, \Delta) = F(R + \Delta) - F(R)$, we have

$$\Psi(R, \Delta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(r, \alpha) r \, dr \, d\alpha$$

(9)

If the range $r$ and the angle $\alpha$ are independent, then $\Psi(R, \Delta)$ is constant with respect to $R$. That is, $\frac{\partial \Psi(R, \Delta)}{\partial R} = 0$. From Equation (9), we derive

$$\frac{\partial F(R + \Delta)}{\partial R} = \frac{\partial F(R)}{\partial R}$$

(10)

Because of the independence of $\alpha$ and $r$, $p(r, \alpha)$ can be factored as

$$p(r, \alpha) = f(r) g(\alpha)$$

(11)

Without loss of generality, we impose that $\int g(\alpha) \, d\alpha = 1$. After factoring, Equation (8) becomes

$$F(R) = \int_{-\infty}^{\infty} f(r) r \, dr$$

Because of the property of the integration, we have

$$\frac{\partial F(R)}{\partial R} = f(R) R$$

(12)

From Equations (10) and (12), we deduce that $f(R + \Delta)(R + \Delta) = f(R) R$. Therefore, $R \left[ \frac{f'(R + \Delta)}{f'(R)} \right] + f'(R + \Delta) = 0$. By making $\Delta$ goes to zero, we obtain $R f'(R) + f(R) = 0$.

The equality $f'(R)/f(R) = -1/R$ can be re-written as $\{\log(f(R))\}' = -\{\log(R)\}'$. After integrating both sides, we obtain

$$\log(f(R)) = -\log(R) + c = \log \left( \frac{1}{R} \right) + c = \log \left( \frac{\xi}{R} \right)$$

Where $c$ is a constant, let $\xi = e'$, we obtain $f(R) = \frac{\xi}{R}$

From Equations (7) and (11), $\xi$ can be calculated and thus $\xi = \frac{1}{R_{\text{max}} - R_{\text{min}}}$

$$f(r) = \frac{1}{(R_{\text{max}} - R_{\text{min}}) r}$$

Therefore, $p(r, \alpha)$ can be re-written as

$$p(r, \alpha) = \frac{1}{(R_{\text{max}} - R_{\text{min}}) r} \times g(\alpha)$$

(13)

If we use a Gaussian function for $g(\alpha)$ with mean $\beta$ and standard deviation $\sigma$, the PDF
\( p(r, \alpha) \) can be re-written as Equation (14). Figure 12 shows the PDF of \( p(r, \alpha) \)

\[
p(r, \alpha) = \frac{1}{(R_{\text{max}} - R_{\text{min}})} \frac{\exp\left[ -\frac{(\alpha - \beta)^2}{2\sigma^2} \right]}{\sqrt{2\pi \sigma}}
\] (14)

Figure 12. The PDF of the landmark position following Equation (14)

5.2 Utilization of the PDF for bearing-only SLAM

We illustrate the application of the PDF for our bearing-only SLAM system. Section 5.2-A describes how the PDF can be updated with a second observation. In Section 5.2-B, we present experimental results on a real robot.

A. Updating the PDF with a second observation

When a second observation is made after a linear motion, the landmark position falls in the uncertainty region which is the intersection of two vision cones rooted at the first and the second observation points \( O_1 \) and \( O_2 \). We denote with \( p_1 \) and \( p_2 \) the PDFs of the landmark positions computed from Equation (14) with respect to \( O_1 \) and \( O_2 \) respectively. Let \( p \) denote the PDF of the landmark position after fusing the sensory readings from \( O_1 \) and \( O_2 \). From the work of Stroupe et al. (2000), we have \( p = p_1 p_2 / \int p_1 p_2 \).
We want to approximate \( p \) with a Gaussian distribution \( q \). To compute the parameters of \( q \), we generate a set \( S \) according to the PDF \( p \) by the Rejection Method (Leydold, 1998). We determine the maximum probability density \( p_{\text{max}} \) of \( p \) by computing \( p(p, p) \) at the intersection of two bearings. The sampling process selects uniformly a sample point \( s \) and a random number \( \nu \in [0, 1] \). If \( p(s)/p_{\text{max}} \) is rejected, otherwise \( s \) is accepted and added to \( S \). The sampling process is repeated until enough points are accepted. Figure 13 shows the generated samples in the uncertainty regions of four landmarks.

The mean \( \mu \) and the covariance matrix \( C \) of \( q \) are obtained by computing the mean and the covariance matrix of \( S \) as previously done by Smith & Cheeseman (1986) and Stroupe et al. (2000). In Figure 13, the contour plots present the PDFs of landmark positions.

The estimated PDFs in Figure 13 are expressed in the robot-based frame \( F_R \). Since the robot is moving over time, its frame changes too. Therefore, it is necessary to change the coordinate systems to express all the estimations in the global frame \( F_L \). We use the method introduced in Section 4 to transfer the samples in \( S \) from \( F_R \) to \( F_L \). After the change of frames, the uncertainties of \( L_1 \) and \( L_2 \) are transferred to other objects. The samples of other objects are taken to approximate the PDFs of the object positions in \( F_L \). Figure 14 shows the distribution of the samples after the change of frames. The contour plots present the PDFs of the object positions in the global frame \( F_L \) associated to the points \((L_1, L_2)\).

![Figure 13](image_url)

Figure 13. The PDFs and the contour plots of four landmarks in the robot-based frame \( F_R \); in this example, the uncertainty region of each landmark is a bounded polygon. The generated samples are distributed in the uncertainty regions.
B. Experimental Results

Our method was evaluated using a Khepera robot equipped with a colour camera (176 x 255 resolution). The Khepera robot has a 6 centimetres diameter. A KheperaSot robot soccer playing field, 105 x 68 square centimetres, was used for the experiment arena (see Figure 15). There were four artificial landmarks in the playing field. The first and second landmarks were placed at the posts of a goal, 30 centimetres apart from each other.

The objective of the experiment is to evaluate the accuracy of the method by estimating the positions of the third and the fourth landmarks. At each iteration, the robot was randomly placed in the field. The robot took a panoramic image and then moved in a straight line and captured a second panoramic image. Landmark bearings were extracted from the panoramic images using colour thresholding.

A total of 40 random movements were performed in this experiment. $R_{\text{min}}$ was set to 3 centimetres, $R_{\text{max}}$ was set to 120 centimetres, and the vision error $\varepsilon$ was ± 3 degrees. Figure 16(b) shows the errors of the estimated landmark positions. For $L_3$, the estimated error was reduced from approximately 9 centimetres at the first iteration to less than 1 centimetre at the last iteration. For $L_4$, the error was reduced from 14 centimetres to 2.77 centimetres. The experiment shows that the estimated error of landmark position is sensitive to the relative distance with respect to $L_1$ and $L_2$. 

Figure 14. After the change of frames from $F_X$ to $F_Y$, the PDFs and the contour plots of $O_1$, $O_2$, $L_1$, and $L_2$ are presented in the global frame $F_G$. 
Figure 15. Experimental setup, the landmarks are the vertical tubes

Figure 16. (a) Diagram of the experimental setup. (b) The errors of the estimated landmark positions

We made another experiment to test the sensitivity of the errors of the landmark positions with respect to the different directions of the robot’s moving trajectories. We let the robot move in four different directions with respect to three landmarks. In Figure 17(a), stars denote the landmark positions and arrows denote the moving trajectories. The robot repeated 10 iterations for each trajectory. The errors on $L_4$ in four trajectories after the tenth iteration were 2.12, 1.17, 1.51, and 13.99 centimetres respectively. The error of the fourth trajectory is large because the robot moves along a line that is close to $L_4$. Therefore, the vision cones at the first and the second observations are nearly identical.
The estimation of the landmark position is more accurate when the intersection of two vision cones is small. This is the case of the second trajectory where the intersection is the smallest among all trajectories.

Figure 17. (a) Trajectories of the robot for the experiment to study the relationship between moving directions and estimated errors of landmark positions. (b) The errors on $L_i$ at each iteration

Although the intersecting area of $L_i$ for the first and the third trajectories are the same, the intersecting areas of $L_1$ and $L_2$ for the first trajectory are much bigger than the areas from the third trajectory. This is the reason why the estimated error from the third trajectory is smaller than the one for the first trajectory.

6. Conclusion

In this chapter, we proposed a vision-based approach to bearing-only SLAM in a 2-dimensional space. We assumed the environment contained several visually distinguishable landmarks. This approach is inspired from techniques used in stereo vision and Structure From Motion. Our landmark initialization method relies solely on the bearing measurements from a single camera. This method does not require information from an odometer or a range sensor. All the object positions can be estimated in a landmark-based frame. The trade-off is that this method requires the robot to be able to move in a straight line for a short while to initialize the landmarks. The proposed method is particularly accurate and useful when the robot can guide itself in a straight line by visually locking on static objects.

Since the method does not rely on odometry and range information, the induced map is up to a scale factor only. In our method, the distance $||L_1 - L_2||$ of two landmarks is taken as the measurement unit of the map. The selection of $L_1$ and $L_2$ is critical for the accuracy of
the map. In Section 3.1, the mathematical derivation shows that the estimated error of a landmark position is proportional to the range of the landmark and the inverse of the relative change in landmark bearings. Choosing \( L_1 \) and \( L_2 \) with larger change in bearings produces a more accurate mapping of the environment.

In the sensitivity analysis, we showed how the uncertainties of the objects’ positions are affected by a change of frames. We determine how an observer attached to a landmark-based frame \( F_L \) can deduce the uncertainties in \( F_L \) from the uncertainties transmitted by an observer attached to the robot-based frame \( F_R \). Each estimate of landmark uncertainties requires a pair of the observations in a straight movement. The simulation in Section 4.3 shows how the uncertainties of landmark positions are refined when the robot moves in a polygonal line.

With dead reckoning, the error of the estimated robot’s location increases with time because of cumulated odometric errors. In our method, we set \( O_1 \) and \( O_2 \) (pair of observation points in a straight movement) at \([0,0]'\) and \([1,0]'\) in \( F_R \). There is no dead reckoning error on \( O_1 \) and \( O_2 \) by construction. In practice, the robot’s movement may not be perfectly straight. However, the non-straight nature of the trajectory can be compensated by increasing the size of the confidence interval of the bearing.

The induced map created by our method can be refined with EKF or PF algorithms. With EKF, the uncertainty region computed from the geometric method can be translated into a Gaussian PDF. With PF, the weights of the samples can be computed with the formulas derived in Section 5.1. Since the samples are drawn from the uncertainty region, the number of samples is minimized.

The accuracy of our method was evaluated with simulations and experiments on a real robot. Experimental results demonstrate the usefulness of this approach for a bearing-only SLAM system. We are currently working on the unknown data association when all landmarks are visually identical. In future work, we will deal with the problems of object occlusion and non-planar environments. That is, the system will be extended from a 2-dimensional to a 3-dimensional space.

7. References


Bearing-only Simultaneous Localization and Mapping for Vision-Based Mobile Robots


Computer Vision is the most important key in developing autonomous navigation systems for interaction with the environment. It also leads us to marvel at the functioning of our own vision system. In this book we have collected the latest applications of vision research from around the world. It contains both the conventional research areas like mobile robot navigation and map building, and more recent applications such as, micro vision, etc. The fist seven chapters contain the newer applications of vision like micro vision, grasping using vision, behavior based perception, inspection of railways and humanitarian demining. The later chapters deal with applications of vision in mobile robot navigation, camera calibration, object detection in vision search, map building, etc.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

http://www.intechopen.com/books/vision_systems_applications/bearing-only_simultaneous_localization_and_mapping_for_vision-based_mobile_robots