Linear 2D Localization and Mapping for Single and Multiple Robot Scenarios

Frank Dellaert[1] and Ashley W. Stroupe[2]
[1]College of Computing, Georgia Institute of Technology, Atlanta, GA
[2]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

Abstract

We show how to recover 2D structure and motion linearly in order to initialize Simultaneous Mapping and Localization (SLAM) for bearings-only measurements and planar motion. The method supplies a good initial estimate of the geometry, even without odometry or in multiple robot scenarios. Hence, it substantially enlarges the scope in which non-linear batch-type SLAM algorithms can be applied. The method is applicable when at least seven landmarks are seen from three different vantage points, whether by one robot that moves over time or by multiple robots that observe a set of common landmarks.

1 Introduction

In many mobile robot applications it is essential to obtain an accurate metric map of a previously unknown environment, and to be able to accurately localize the robot(s) within it. The process of reconstructing such a map from odometry and sensor measurements collected by one or more robots is known as Simultaneous Localization and Mapping (SLAM) [1]. Sensors that are commonly brought to bear on this task include cameras, sonar and laser range finders, radar, and GPS.

In the case a single robot observes a set of landmarks, SLAM algorithms are often based on variable dimension Kalman filters [2, 1, 3]. In these on-line approaches, the robot pose and the landmarks positions are recursively updated as each new measurement is obtained. Because of the recursive nature of these algorithms, and because odometry can be used to predict changes in pose and relative landmark positions, a good initial estimate is always available in the Kalman update step. This is essential in practice, as the measurement equations involved are often non-linear in the state-variables.

However, if *multiple* robots observe a common set of landmarks, a recursive algorithm is not applicable. While the SLAM problem can be formulated in terms of non-linear minimization, an important problem in practice is the existence of local minima which can prevent the optimization process from converging. In addition, even if they converge to the correct solution, this is often slow when the initial estimate is far from the global minimum. Thus, an essential element to solving the SLAM problem in the multiple robot scenario is the ability to easily obtain a good initial estimate for the solution.

We show that, with planar motion and bearings-only measurements, an initial estimate is efficiently provided by a *linear* algorithm borrowed from computer vision. To this end, the bearings are converted to projective coordinates in a virtual 1-D camera, after which a linear 2D method for projective structure from motion (SFM) is used to recover the position of the landmarks and the robot poses. As we start from bearing measurements, no calibration is needed and a *metric* reconstruction is obtained up to a 2D similarity transform.

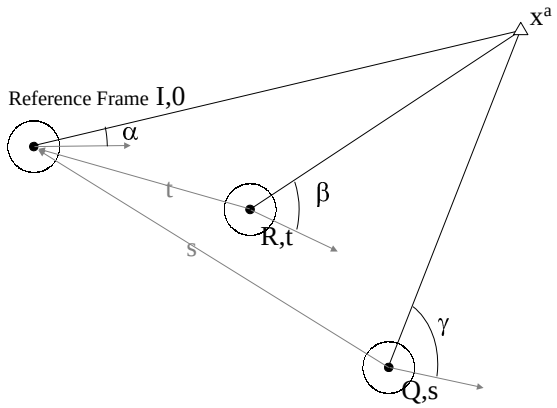


Figure 1: The SLAM problem is to recover both the positions x^a of the landmarks and the relative motion parameters R, t and Q, s for the second and third robot, respectively, from bearing measurements.

Whereas the underlying mathematics has been presented before in the computer vision community, the present paper provides a synthesis of results spread over several relatively inaccessible papers. We also show how to convert the bearings-only SLAM problem to the 2D projective SFM problem, and omit details irrelevant to the SLAM problem (such as self-calibration). We believe the result to be of considerable practical value to researchers interested in the bearings-only SLAM problem. The resulting linear algorithm provides a quick approximate solution to the SLAM problem, which is useful to

1. avoid local minima of non-linear minimization,
2. save computation by starting near the global minimum,
3. provide a quick and dirty SLAM estimate in case non-linear minimization is infeasible.

The technique is applicable for a single robot or in the case where some or all of the robots make multiple observations over time. In this case to it can serve to initialize an incremental estimator, e.g. a variable dimension Kalman filter [2, 3].

2 Simultaneous Localization and Mapping

2.1 Problem Formulation

Assume an unknown environment is observed by one or multiple robots with m different poses a, b, c, etc., and n landmarks x^a are observed (see Fig. 1). The superscript a indicates

that landmark coordinates are expressed with respect to the *reference pose a*. The bearings-only SLAM problem can be stated as follows: given *m* bearing measurements $\alpha, \beta, \gamma, \ldots$ for each of the *n* landmarks, recover the *n* landmark positions and the *m* robot poses. Note that the solution can only be recovered up to a similarity transform in the plane.

2.2 Maximum a Posteriori SLAM

If measurements are noisy, the problem is best formulated as *maximum a posteriori* (MAP) estimation. The unknowns are the $3m$ motion parameters M and the $2n$ landmark positions X. The data Z consists of mn bearing measurements. The MAP estimate is the set of parameters $\{M,X\}^*$ that maximizes the posterior probability $P(M,X|Z)$ of M,X given the data Z:

$$\begin{aligned}\{M,X\}^* &= \text{argmax}_{M,X}\, P(M,X|Z) \\ &= \text{argmax}_{M,X}\, P(Z|M,X)P(M,X)\end{aligned}$$

The posterior probability $P(M,X|Z)$ is the product of the *likelihood* $P(Z|M,X)$ and a prior $P(M,X)$ on the landmark and motion parameters. In typical SLAM scenarios there is no (strong) prior information on the position of landmarks, but odometry provides a prior on the motion parameters of individual robots.

To obtain the MAP estimate, a *measurement model* is needed. The bearing measurement α_{ij} taken by the i^{th} robot on landmark x_j^a can be predicted by first transforming the 2D landmark position into the i^{th} coordinate frame and then taking the arc-tangent:

$$\alpha_{ij} = atan2(R_i x_j^a + t_i) + n_{ij}$$

where R_i and t_i are the rotation and translation, respectively, of the reference frame with respect to the i^{th} frame, and n_{ij} is measurement noise.

2.3 SLAM as Non-linear Least-Squares

If we assume independent normally distributed noise n_{ij} on the bearing measurements, then the likelihood $P(Z|M,X)$ can be factored as follows:

$$P(Z|M,X) = \prod_{i,j} \mathcal{N}(\alpha_{ij}; atan2(R_i x_j^a + t_i), \sigma^2)$$

where $\mathcal{N}(z;\mu,\sigma^2)$ is the normal density with mean μ and variance σ^2. Thus, if no prior is available, the *maximum likelihood* (ML) estimate can be found by minimizing the following non-linear least squares criterion:

$$\{M,X\}^* = \underset{M,X}{\text{argmin}} \sum_{i,j} (\alpha_{ij} - atan2(R_i x_j^a + t_i))^2$$

2.4 Number of Measurements Needed

It is of interest how many landmarks are needed to obtain a solution in the general case. The degrees of freedom of the system are the number of measurements minus the number of parameters, plus 4 because the solution can only be recovered up to a similarity transform in the plane (described by 4 parameters). Thus:

$$DOF = mn + 4 - (3m + 2n)$$

First, note that *with only two views, the system can not be solved*. This can be seen by setting $m = 2$, in which case we get $DOF = -2$. This is also intuitively clear: we can place the robots anywhere and always get a feasible solution for any number of landmarks by intersecting the lines of sight for each landmark.

With three views, the minimal number of landmarks is 5. Indeed, setting $m = 3$ we get $DOF = n - 5$.

2.5 The Problem of Local Minima

The problem with the non-linear method outlined above is that is sensitive to local minima. The nonlinear least-squares minimization proceeds iteratively, and we are only guaranteed to find the globally correct solution if we start from an initial estimate that is in the *basin of attraction* of the global minimum. If this is not the case, the iterative minimization procedure will get stuck in a local minimum.

This is an important problem in practice, as simulation experiments with randomly generated problems show that typically more than 50% of all runs end up in a local minimum. The more views, the more important the problem becomes.

In the case of a single robot, odometry can be used to provide a good initial estimate. This is the basis of existing SLAM methods that typically use variable dimension Kalman filters. However, in the multiple robot case or if no odometry is available there is currently no good solution.

3 A Linear Solution

In this section we describe a linear solution to the bearings-only SLAM problem that does not suffer from local minima, but immediately finds the globally optimal solution. This provides a basis to solve the SLAM problem in the multiple robot case or in the single robot case when there is no odometry available.

The linear method is based on the linear *structure from motion* (SFM) algorithms developed over the last few years in the computer vision community [4], but specialized to the 2D, bearings-only case [5]. We can easily transform the bearings-only problem into a 2D SFM problem, by converting bearing measurements to 1D *image measurements* in a virtual camera, and solving the associated 2D SFM problem linearly. The resulting solution can then be fine-tuned by non-linear minimization, if desired.

3.1 Converting to a Projective Formulation

Below we introduce homogeneous coordinates, as the linear methods are based on projective geometry. Bearings measurements taken by the robots will be converted to measurements in a set of 1D perspective cameras or *views*, one for each robot. The views are denoted by Ψ_1, Ψ_2, Ψ_3, etc. 1D image measurements are given by their projective coordinates $u^A \triangleq (u^1\, u^2)$, $v^B \triangleq (v^1\, v^2)$, etc..., where the uppercase superscripts *A, B,...* indicate in which view the measurements where taken. The landmarks are described by 2D projective coordinates $x^a \triangleq (xyz)^T$. The superscript a in x^a refers to the reference view, which we arbitrarily take to be the first view Ψ_1.

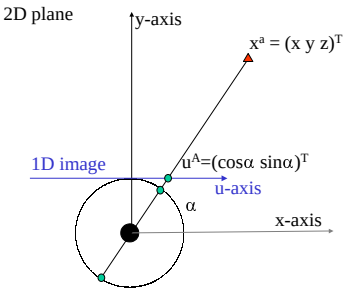


Figure 2: Converting from bearings to homogeneous image measurements. With the conversion $u^A \leftarrow (\cos\alpha \sin\alpha)^T$ a virtual camera is created as shown, parallel to the x-axis at $v = 1$.

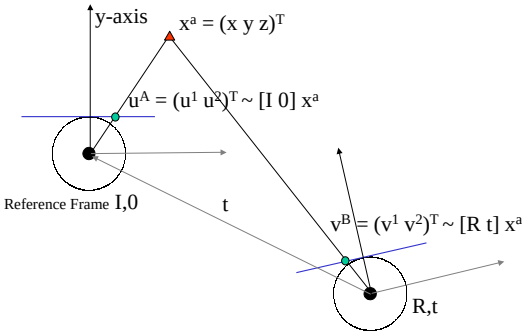


Figure 3: 2D to 1D projection in homogeneous coordinates is a linear operation. Above we project the 2D point $x^a \triangleq (xyz)^T$ into the reference frame Ψ_1 and into a second view Ψ_2 with motion parameters R, t.

2D coordinates in the other views take superscripts b, c, etc. Homogeneous coordinates are only defined up to a scale, i.e. $x^a \equiv y^a$ iff $x^a = \lambda y^a$. The 1D and 2D projective spaces will be referred to as $\mathbb{P}^1$ and $\mathbb{P}^2$, respectively.

We can easily convert bearing measurements α to 1D image measurements u^A in a *virtual camera* as follows:

$$u^A \leftarrow (\cos\alpha \sin\alpha)^T$$

The virtual camera is located at $y = 1$ on the y-axis, parallel to the x-axis, as illustrated in Fig.2. The boundary cases of a bearing measurement with $\alpha = 0$ or $\alpha = \pi$ is handled automatically, as in both cases the 1D projective coordinate will be $(1\,0)^T$, i.e. the "point at infinity". Note that as there is only one virtual camera for $y > 0$, the mapping above is a 2-to-1 mapping: bearings that differ by π are identified.

In homogeneous coordinates, the projection from 2D to 1D is a linear operation. This is illustrated for one 2D landmark and two views in Fig. 3. In the reference view Ψ_1, which we can arbitrarily place at the origin, we have

$$u^A \triangleq (u^1\, u^2)^T \equiv [I\,0]\, x^a$$

and in a second view Ψ_2 we have

$$v^B \triangleq (v^1\, v^2)^T \equiv [R\,t]\, x^a$$

with R and t the rotation and translation of view Ψ_2.

3.2 Recovering Landmarks

Consider first the sub-problem of recovering the landmarks in the case that the robot poses are known. As discussed above, a landmark x^a gives rise to an image measurement in each view given by

$$v^B \triangleq (v^1\, v^2)^T \equiv [R\,t]\, x^a$$

which is equivalent to $v^B \times [R\,t]\, x^a = 0$. Written out explicitly this yields the homogeneous equation

$$(v^1 R_{21} - v^2 R_{11})x + (v^1 R_{22} - v^2 R_{12})y + (v^1 t^2 - v^2 t^1)z = 0$$

where we assumed $x^a = (xyz)^T$. Each view gives one such equation, and hence given at least 2 views we can linearly recover $(xyz)^T$. This makes intuitive sense: the landmark can be recovered by simply intersecting the viewing rays.

In practice this is done using singular value decomposition (SVD). To this end, we first form a $m \times 3$ data matrix D, where each row is formed by the three coefficients of the equation in the corresponding view. For example, for three views D is equal to

$$\begin{bmatrix} -u^2 & u^1 & 0 \\ v^1 R_{21} - v^2 R_{11} & v^1 R_{22} - v^2 R_{12} & v^1 t^2 - v^2 t^1 \\ w^1 Q_{21} - w^2 Q_{11} & w^1 Q_{22} - w^2 Q_{12} & w^1 s^2 - w^2 s^1 \end{bmatrix}$$

where Q and s are the rotation and translation parameters of the third view Ψ_3, respectively, and the image measurement in Ψ_3 is given by $w^C \triangleq (w^1\, w^{2)^T}$.

The matrix D is then decomposed using SVD:

$$D_{m\times 3} = U_{m\times 3}\, \Lambda_{3\times 3}\, V_{3\times 3}$$

where the columns of $V_{3\times 3}$ contain the eigenvectors e_i of $D^T D$. The eigenvector e^* corresponding to the minimum eigenvalue λ^* minimizes the sum of squares of the residual, subject to $||e^*|| = 1$. The homogeneous coordinate of the recovered landmark is thus $x^a \equiv e^*$.

3.3 Recovering the Motion

A second sub-problem is recovering the relative motion parameters in the case that the *epipoles* are known. The epipole is simply the projection in one view of the center of projection of a second view. Finding the epipoles is at the core of the linear approach and is discussed in the next section, Section 3.4.

If the second view Ψ_2 has relative motion parameters R and t, respectively, then the epipole in Ψ_2 is $t \triangleq (t^1\, t^2)^T$. The translation between Ψ_1 and Ψ_2 is only defined up to a scale and its direction is given directly by the epipole t. The situation is illustrated in Fig. 4. The bearing α from Ψ_2 to Ψ_1 satisfies

$$(\cos\alpha \sin\alpha) \equiv t$$

Let us denote the epipole in the first view Ψ_1 as $e^A = (e^1\, e^2)^T$. The bearing β to the second view Ψ_2 satisfies

$$(\cos\beta \sin\beta) \equiv e$$

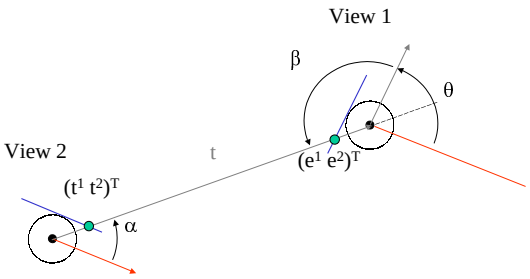


Figure 4: Determining the relative orientation of two views from the epipoles t and e. The angles α and β are recovered directly, up to π radians. The relative orientation θ is then $\alpha - \beta$.

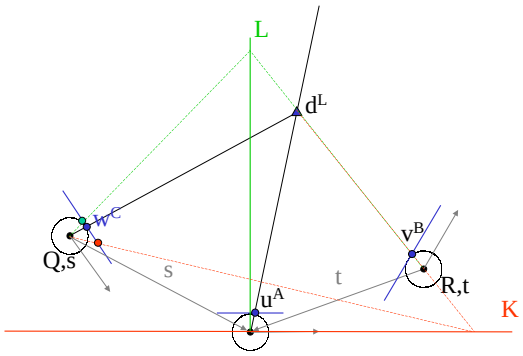


Figure 5: Intrinsic homographies K and L, see text.

From these two relations, we get two solutions for the relative orientation $\theta = \alpha - \beta$:

$$\begin{cases} \theta = atan2(t) - atan2(e) \\ \theta = atan2(t) - atan2(e) + \pi \end{cases}$$

Below we are interested particularly in the three view case. For three views and known epipoles we can use the method above to recover the relative orientations θ and γ of view Ψ_2 and Ψ_3 with respect to the reference view Ψ_1. By fixing the scale of the translation of view Ψ_2, i.e. choosing a specific scale for t, we can recover the location of the third view by triangulation, using the epipoles of view Ψ_3 in view Ψ_1 and Ψ_2 as image measurements.

3.4 A Linear Method for Three Views

In this section we describe the linear method of recovering 2D structure and motion for three views. The exposition below is a synthesis of material found in [5, 6, 7, 4]. For three views and at least 7 landmarks, there exists a linear algorithm to recover a set of coefficients that completely describes the geometry of the three views. From these coefficients we can recover all 6 epipoles in the three views. Then, as described above, from the epipoles we can recover the relative motion and subsequently the position of the landmarks.

3.4.1 The Trifocal Tensor: The intuition underlying the linear algorithm is simple. A triple of corresponding measurements in three views cannot be independent: if two measurements are given, the location of the third measurement can be predicted simply by triangulating the first two measurements in Ψ_1 and Ψ_2 and re-projecting in the third view Ψ_3. It is well known that this three-view relationship can be expressed as a *trilinear constraint* of the form

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} T_{ijk}\, u^i v^j w^k = 0 \qquad (1)$$

where $u^A \triangleq (u^1\, u^2)$, $v^B \triangleq (v^1\, v^2)$, and $w^C \triangleq (w^1\, w^2)$ are the image measurements in the three views, respectively, and the T_{ijk} are the 8 *trifocal tensor coefficients*. They are so called because they can be arranged in a $2 \times 2 \times 2$ tensor, the *2D trifocal tensor*. Together, they completely describe the relative geometry of the three 1-D virtual cameras to which the bearing measurements are converted.

While there are 8 trifocal tensor coefficients, the trifocal tensor is only defined up to a scale and as such has only 7 degrees of freedom.

3.4.2 Recovering the Trifocal Tensor: We can linearly recover the tensor coefficients T_{ijk} using SVD, in the same way we recovered the landmark positions. Indeed, each constraint of the form (1) contributes one homogeneous equation on the coefficients T_{ijk}. To recover 8 coefficients up to a scale, we need at least 7 equations, hence the requirement of having at least 7 landmarks.

The algorithm is the same as the one in Section 3.2, except each row in the D-matrix is now formed by the entries $u^i v^j w^k$ from equation 1.

3.4.3 Recovering Epipoles: To recover the epipoles from the trifocal tensor, we use a technique due to Shashua for the 3D case [7]. This necessitates a geometric interpretation of the trifocal constraint in terms of homographies. A *homography* is a mapping between projective coordinates on two lines, induced by a third line. For example, if we take the line of sight through an image coordinate u^A in view Ψ_1, a homography H_B^C between the image coordinates in view Ψ_2 and Ψ_3 is induced:

$$w^C \equiv H_B^C v^B$$

where H_B^C is 2×2. The subscript B and superscript C indicate that H_B^C goes from view Ψ_2 to view Ψ_3.

Two special homographies, the *intrinsic homographies* K and L, are induced by taking u^A to be $(1\,0)^T$ and $(0\,1)^T$, respectively. As shown in Fig. 5, the intrinsic homography K is induced by the line through the optical center of view Ψ_1 and parallel to its image plane. The intrinsic homography L is induced by the line perpendicular to that, through $u^A = (0\,1)^T$. The coefficients of K and L can be directly obtained from the trifocal tensor, as each homography is simply the two-view constraint obtained from (1) by filling in u^A:

$$\sum_{j=1}^{2}\sum_{k=1}^{2} T_{1jk}\, v^j w^k = 0$$

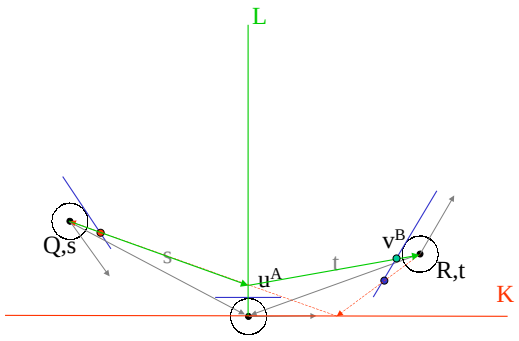


Figure 6: A special homography from the second view Ψ_2 to itself is defined by first mapping through K, and then back through L, i.e. $M \equiv L^{-1}K$.

$$\sum_{j=1}^{2}\sum_{k=1}^{2} T_{2jk} v^j w^k = 0$$

from which get the homography matrices (from Ψ_2 to Ψ_3):

$$K_B^C = \begin{bmatrix} -T_{112} & -T_{122} \\ T_{111} & T_{121} \end{bmatrix}$$

$$L_B^C = \begin{bmatrix} -T_{212} & -T_{222} \\ T_{211} & T_{221} \end{bmatrix}$$

Different arrangements of the trifocal coefficients give intrinsic homographies between all ordered pairs of the three views.

The epipoles are now found as follows. We can define a homography M from view Ψ_2 to itself, by mapping a point v^B to Ψ_3 through K, and then back to Ψ_2 by means of L^{-1}, as illustrated in Fig. 6. The 2×2 homography matrix is then $M \equiv L^{-1}K$. The only points that are mapped to themselves under this mapping M are the epipoles, i.e. *the epipoles are the eigenvectors of M*. Thus, by performing an eigenvalue decomposition of M (which can be done in closed form for 2×2 matrices) we obtain the two epipoles e_1^B and e_3^B in view Ψ_2. The superscript B refers to the fact that they are image points in view Ψ_2, while the subscript 1 or 3 refers to the optical center it is the image of (that of view Ψ_1 or view Ψ_3). The corresponding epipoles in view Ψ_1 and Ψ_3 can be obtained by pushing the epipoles e_1^B and e_3^B through either K or L (using K_B^A for view Ψ_1 and K_B^C for Ψ_3), as epipoles are always mapped to each other (by any homography !).

3.5 Summary

The linear algorithm to recover structure and motion for the three view case is now summarized:

Step 1. Recover the trifocal tensor coefficients T_{ijk} form at least 7 three-view correspondences using the '7-point algorithm' from Section 3.4.2.

Step 2. Recover the epipoles in view Ψ_2 as the eigenvectors of $M \equiv L^{-1}K$, and the corresponding epipoles in views Ψ_1 and Ψ_3 by applying the homographies K_B^A and K_B^C (Section 3.4.3).

Step 3. From the epipoles, recover the relative motion parameters R,t and Q,s (Section 3.3).

Step 4. Triangulate the position of the landmarks through SVD (Section 3.2).

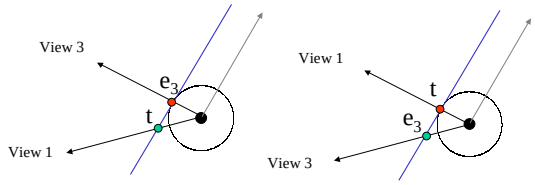


Figure 7: The two recovered epipoles e and t in view Ψ_2, obtained as the eigenvectors of M, can be assigned in two ways to the view Ψ_1 and Ψ_3. Both possibilities lead to self-consistent structure and motion solutions in the three-view case, regardless of the number of landmarks involved.

3.6 Multiplicity of Solution

There are, however, two ways in which we can assign the eigenvectors of M to the epipoles in step 2. This is illustrated in Fig. 7. Both choices lead to a different set of 6 epipoles. Remarkably, *both choices lead to a completely self-consistent solution for structure and motion*. This is a fundamental ambiguity of the three view case, and is true for any number of landmarks, counter-intuitive though it seems. Thus, the algorithm will always output two valid reconstructions consistent with the image measurements in the three views.

When starting from bearing measurements, we can frequently pick the correct solution from the two possible structure from motion solutions. Recall that, in order to obtain a linear algorithm, we convert the bearing measurements α to image measurements u^A by

$$u^A \leftarrow (\cos\alpha \sin\alpha)^T$$

but we lose some information in the process. In particular, we lose the distinction between bearings α and $\alpha+\pi$. After we have recovered the structure and motion, however, we can re-calculate the actual bearings and check whether they agree with the measured bearings. Frequently, one of the two solutions will contain landmarks that are inconsistent (i.e. differ by π) with the measured bearings, while the other solution is consistent. Thus, by checking this, we can frequently choose correctly between the two solutions.

Unfortunately, even with bearing measurements it is possible to obtain *two* consistent solutions. In this case, the only way to disambiguate between the two three-view solutions is by adding a fourth view. To do this, we simply add the view to both solutions (see below) and look at the SVD residual. The solution with the lowest residual is then chosen as the correct one. Adding views is described in the next section.

3.7 More Than Three Views

A linear method that treats all views simultaneously is not available, but given an initial three-view geometry and its recovered structure we can easily recover the relative motion of additional views. This can be done similarly to recovering the position of the landmarks as described in Section 3.2.

For example, when adding a fourth view Ψ_4, each measurement u^D in Ψ_4 yields a homogeneous equation in the motion

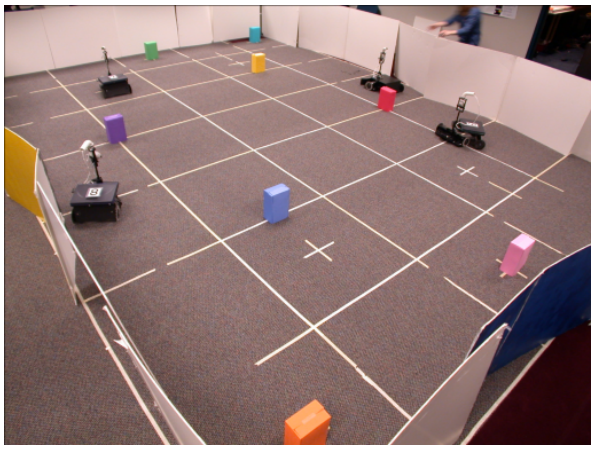


Figure 8: Experimental setup with Minnow robots and colored landmarks. The picture corresponds to the arrangement of Fig. 9 below (although one landmark is missing in the picture).

parameters of Ψ_4. As before, let us denote these parameters by R and t, where $t \triangleq (t^1\, t^2)^T$ and

$$R \triangleq \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

with $c = \cos\theta$ and s=$\sin\theta$ for some θ. Now, any landmark $x^a \triangleq (xyz)^T$ is projected into Ψ_4 according to

$$u^D \equiv [Rt]x^a$$

Written out as a cross product, this yields the homogeneous equation below in the motion parameters t^1, t^2, c, and s:

$$(yu^1 - xu^2)c + (xu^1 + yu^2)s + (-zu^2)t^1 + (zu^1)t^2 = 0$$

Using the by now familiar method of SVD, we can recover the motion parameters up to a scale given at least three common measurements with the recovered structure so far. After performing SVD and taking the eigenvector corresponding to the smallest eigenvalue, the correct scale can be recovered by imposing the constraint $c^2 + s^2 = 1$.

4 Results

4.1 Experimental Results

In terms of experimental validation, we performed two experiments with a team of mobile robots. The experimental setup is illustrated in Fig. 8. As hardware platform we used a team of 4 "Minnow" robots, a class of small mobile robots developed at CMU and based on the Cye robot, an inexpensive and commercially available platform. As landmarks we used cardboard boxes (about 16" wide, 10" deep, 20" tall) covered with colored construction paper. The experimental area was about 5x7 meters, enclosed by white poster-board walls.

A commercial USB camera provides sensory input in the form of images at a resolution of 360×240 pixels. Landmarks were detected using blob detection and identified through color analysis. The cameras were calibrated so that yaw angle is easily calculated from the position of the color blob in the image. Because the cameras have a limited field of view, 4 images are taken by each robot in each direction, providing a 360 degree field of view of the environment.

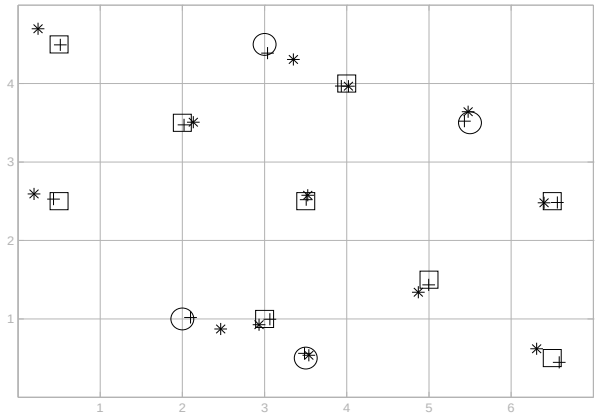


Figure 9: Multiple robot setup: first experiment.

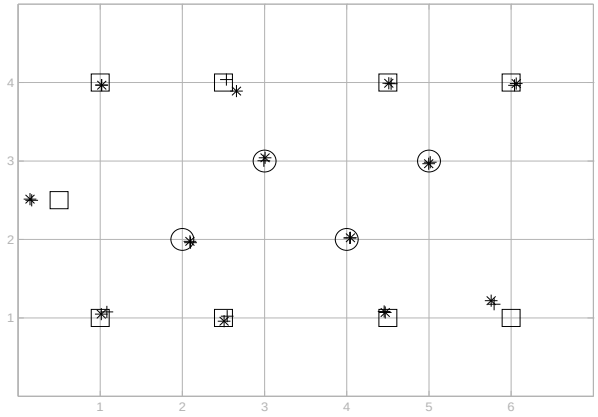
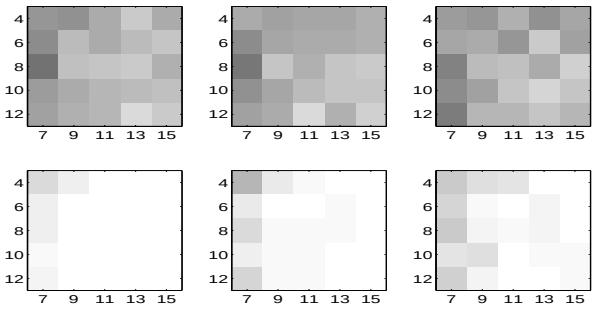


Figure 10: Multiple robot setup: second experiment.

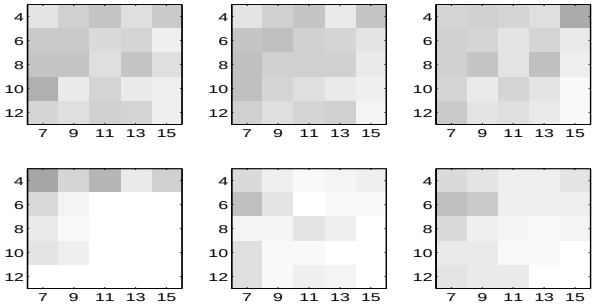
The first experiment corresponds to the arrangement in Fig. 8, and is illustrated in Fig. 9. The ground truth locations are depicted as circles for the robots and squares for the landmarks. The standard deviation of the error on the bearings obtained by the vision system was 2.23 degrees.

The result of applying the linear method is shown in Fig. 9 as asterisks, which represent the recovered position of both the robots and the landmarks. Because these are only recovered up to a 2D similarity transform, they were first optimally aligned with the ground truth to make comparison possible. Note that the alignment process simply recovers the 4 unknown ambiguities and does not improve or degrade the results of the linear step. As can be seen from the figure, the recovered position of two of the robots and some of the landmarks has appreciable error. However, the reconstruction is good enough to ensure fast convergence of a subsequent non-linear refinement step, the result of which is shown using the '+' symbols.

In a second experiment, illustrated in Fig. 10, we created a setup where the robot-team is surrounded by the landmarks. Such a situation would occur when using landmarks on the horizon (mountain tops, large buildings), or when all the robots are within the same open space and landmarks are



(a) Mixed config.: 0.1, 0.5, and 1 degree stddev.



(b) Enclosed config.: 0.1, 0.5, and 1 degree stddev.

Figure 11: Convergence results for synthetic data for (a) mixed, and (b) enclosed configuration. In each case, 50 synthetic datasets were generated for 25 different combinations of *m* and *n*, i.e. 1250 data sets, and this with three different noise levels, for a total of 3750 data sets. The images show the percentage of cases in which SLAM failed to converge without (top rows) and with linear initialization (bottom rows). White=0%, black=100% failures.

available only on the perimeter of the space. The standard deviation of the bearing error in this case was 1.53 degrees. The linear and subsequent non-linear reconstruction are again shown as asterisks and pluses in Fig. 10.

In both experiments, the linear reconstruction provided a good initial estimate for the solution obtained by subsequent non-linear minimization. Of course, both are expected to be different from the ground truth because of measurement error. As expected, this difference is more pronounced in the first experiment due to the larger bearings errors.

4.2 Qualitative Analysis

In order to characterize the behavior of the linear method under different circumstances, we ran it on a large number of synthetic data-sets. We randomly generated two types of configurations: (a) mixed, i.e. landmarks share the same area as the robots, as in Fig. 8 and 9, and (b) enclosed, i.e. landmarks enclose the robot team, as in Fig. 10. For each type, we generated 50 synthetic data-sets for each of 25 different combinations of *m* (#robots ranging from 4 to 12) and *n* (#landmarks ranging from 7 to 15). Measurements were obtained by taking the ground truth bearings and adding Gaussian noise with standard deviations 0.1, 0.5, and 1 degrees.

The results are best appreciated graphically, as shown in Fig. 11. In this figure we graphically show in what percentage of the cases the non-linear minimization process fails to converge (i) using a random initialization (top three images in Fig. 11a and 11b), and (ii) using the linear method to obtain an initial estimate (bottom three images in Fig. 11a and 11b).

The results show that the linear method substantially increases the number of cases in which SLAM converges to the global minimum. However, it is sensitive to measurement noise, particularly in near-minimal configurations. The sensitivity to noise is most noticeable in the "enclosed" configurations. The noise-sensitivity decreases for both types as more landmarks and/or robots are added. For errors in the range 0-0.3 degrees, the linear method almost always leads to convergence, with rare exceptions in the minimal configuration cases. We conjecture that those cases are the result of having randomly generated near-degenerate data-sets (i.e. the landmarks or robots are not in general configuration).

5 Conclusion

Linear methods for projective structure recovery can be successfully applied to the bearings-only SLAM problem under the assumption of planar motion. This is an important case in practice, and we hope that this new tool will ease the application of batch-type SLAM methods to multiple robot scenarios.

Acknowledgments

We wish to acknowledge DARPA's Mobile Autonomous Robot Software Program and the Northrop Grumman Corporation for their support of this research, as well as Tucker Balch for providing the Minnow robots for our experiments.

References

[1] J. Leonard and H. Durrant-Whyte, "Simultaneous map building and localization for an autonomous mobile robot," in *IEEE Int. Workshop on Intelligent Robots and Systems*, pp. 1442–1447, 1991.

[2] I. Cox, "Blanche—an experiment in guidance and navigation of an autonomous robot vehicle," *IEEE Trans. on Robotics and Automation*, vol. 7, no. 2, pp. 193–204, 1991.

[3] J. Castellanos and J. Tardos, *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach*. Boston, MA: Kluwer Academic Publishers, 2000.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[5] L. Quan, "Uncalibrated 1D camera and 3D affine reconstruction of lines," in *CVPR97*, pp. 60–65, 1997.

[6] L. Quan and T. Kanade, "Affine structure from line correspondences with uncalibrated affine cameras," *PAMI*, vol. 19, pp. 834–845, August 1997.

[7] A. Shashua and M. Werman, "Trilinearity of three perspective views and its associated tensor," in *Int. Conf. on Computer Vision (ICCV)*, pp. 920–925, 1995.