Proceedings of the 1998 IEEE
International Conference on Robotics & Automation
Leuven, Belgium • May 1998

# Two-Dimensional Landmark-based Position Estimation from a Single Image[1]

Antonio J. Muñoz and Javier Gonzalez

Departamento de Ingenieria de Sistemas y Automatica, Universidad de Malaga
Campus Teatinos, 29080 Malaga, SPAIN. E-mail: ajmunoz@ctima.uma.es
E-mail: jgonzalez@ctima.uma.es

## Abstract

*This paper addresses the problem of self-location for a mobile robot equipped with a single camera moving in an indoor environment. The robot is provided with a two-dimensional map where the position and some attributes of landmark points are stored. The proposed algorithm first determines the observation rays of vertical edges extracted from one image and then finds an interpretation for these rays in terms of the landmark points. This interpretation is driven by a set-based approach that compels the actual pose to lie in a solution region and not to violate the landmark attributes. Based on the ray-landmark matches provided by the selected interpretation, an optimization procedure is used to come up with the pose for which the mean square angular error is minimum. Finally, we present experimental results that demonstrate the performance of the system.*

## 1. Introduction

The problem of self-location for a mobile robot has been studied by many researchers and a variety of techniques have been proposed for solving it. These techniques vary significantly depending on the type of environment in which the mobile robot has to navigate, on the prior knowledge of the environment, on the task to realize and, of course, on the type of sensor used by the robot [1].

At the University of Malaga we are investigating the problem of estimating the pose (position and orientation) for a mobile robot equipped with a single camera moving in an indoor environment. We assume that the robot has a two-dimensional map where the positions and some attributes of vertical landmarks (vertically oriented parts of fixed objects such as doors, wall junctions, windows, etc.) are stored. The problem can be divided into three steps: 1) detect features in the image that correspond to vertical landmarks of the environment, 2) find the best interpretation of the directions (rays) to these features that is consistent with the landmark attributes, and 3) compute a pose from those matches. In this paper we deal with the two latter problems.

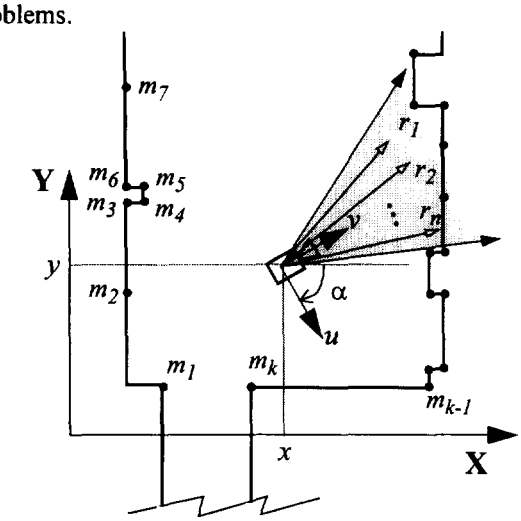



Figure 1.- The landmark-based pose estimation problem.

We believe that vertical landmarks are attractive because they appear very frequently in a typical indoor environment and they can be easily and reliably extracted from the image. Provided that the camera is mounted parallel to the floor, vertical landmarks of the environment will project as vertical features onto the image plane and, therefore, the correspondence problem can be formulated in the two-dimensional plane (as shown in figure 1).

Similar approaches have been proposed by Sugihara [2] and Krotkov [3]. However, they do not deal with the effects of observing *false positive* (detecting features that do not correspond to a known landmark). Another related work is the one reported by Sutherland and Thompson [4] for unstructured environments, where the landmarks are mountain peaks of a topographic map. They analyze how

1. This work has been supported by the Spanish Government under the CICYT project TAP96-0763.

measurement errors affect errors in the localization and propose a simple algorithm to exploit the geometric properties of landmark in the environment in order to decrease errors in the localization. The results presented by all of them are limited to simulated data.

Unlike the above referred works, the algorithm we present here includes some attributes for the landmarks in order to efficiently search the tree and improve the robustness of the algorithm, ruling out the possibility of erroneous matching. Also, a rough estimate of the robot location provided by the dead-reckoning system is considered to simplify the search by predicting which landmarks could match with each ray.

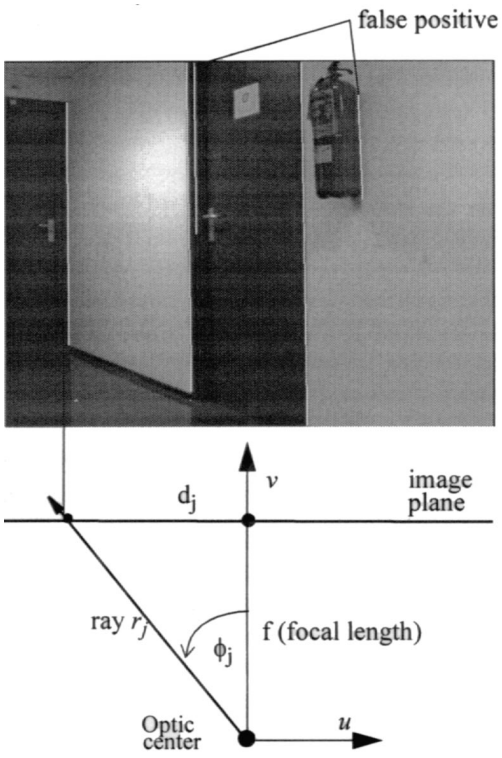



Figure 2.- Extracting rays from vertical edges of an image by using the pin-hole camera model.

## 2. Problem Statement

Let $(x, y, \alpha)$ be the triplet that defines the local robot coordinate system $\{u, v\}$ with respect to the world coordinate system $W$ (see figure 1a).

Let $M=\{m_1, ..., m_k\}$ be the set of landmark points stored in the map and expressed in the world coordinate system $W$. Let $R=\{r_1, ..., r_n\}$ be the set of rays extracted from an image taken at the robot pose $(x, y, \alpha)$.

Intuitively, a ray $r_i$ should correspond to a landmark $m_j$, denoted by the pair $(r_i, m_j)$, if $r_i$ is the ray obtained for the vertical edge which is the projection of $m_j$ onto the image. Since the rays are not distinguishable from each other, we can not establish such a one-to-one correspondence between rays and landmarks. Instead the problem has to be formulated globally by looking for an *admissible interpretation* of the whole set of rays $R$.

In the ideal case when rays and landmark points from the map are error-free, an interpretation is said to be admissible if there exist a pose from where each ray of $R$ pierces the landmark point specified by the interpretation. Notice that in this case both problems correspondence and pose estimation are solved simultaneously as an elementary geometric problem (see [2] for more detail). In practice, since errors are inevitably present in the rays[1] (we assume that the landmarks are error-free), the above criterion is no longer valid and therefore it is necessary to introduce some uncertainty model. In addition, an interpretation should account for rays that have no-correspondence to any known landmark (false positive) as well as for the opposite, that is, landmarks of the map that have not been detected in the image. Obviously, in these cases, the correspondence problem result in a more difficult task that suffers from the ambiguity problem and requires a reliable modeling of the errors involved. In particular, our approach considers that the errors in the observation are within a tolerance. This leads to a set-based algorithm that combines information via intersection of different uncertainty regions. Of this kind is also the work reported by Atiya and Hager for localizing a mobile robot [5].

According to the set-based approach if two regions A and B are known to contain the actual pose of the robot, then $A \cap B$ must also contain that pose. Consequently, each new region that we incorporate to the problem actuates as an additional constraint that provides a tighter bound of the unknown pose.

The first constraint region comes from the uncertainty in the dead-reckoning pose $P_0=(x_0, y_0, \alpha_0)$. We model this uncertainty as a circle $C_0$ of radio $a$, centered at $(x_0, y_0)$. Similarly, a maximum angular error of $\pm\alpha_{error}$ is considered for the orientation $\alpha_0$. This uncertainty region is considered to be proportional to the distance traversed by the robot.

Additional constraint regions come from the fact that the viewpoint from where two landmarks are observed with a given angle must lie on a circumference [2, 3]. Since the error in the rays is within a tolerance, the viewpoint is restricted to lie on a thickened ring as shown in figure 3a. Next, we describe this in more detail.

1. Mostly, these errors are caused by the imperfect localization of the vertical edges in the image as well as errors introduced by the camera model being used.

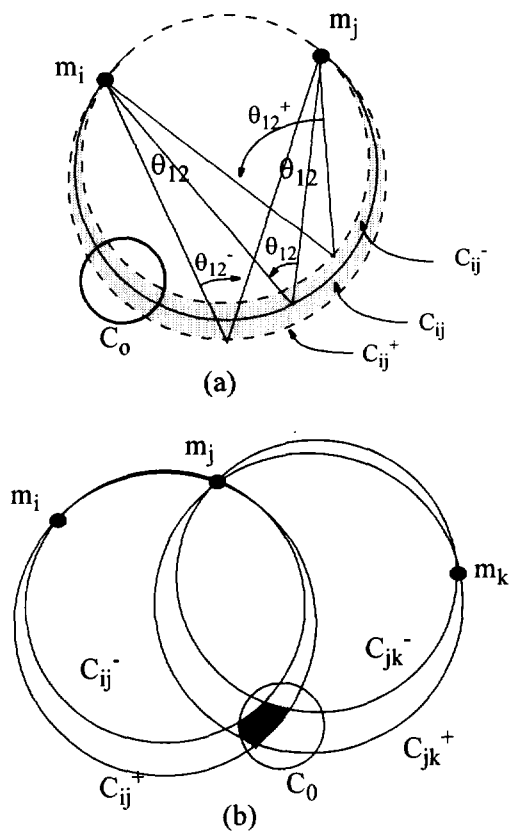



Figure 3.- (a) The error between two rays constraints the viewpoint to a thickened ring. (b) A typical uncertainty region when three pairs ray-landmark are established.

Let $\theta_{12} = \phi_1 - \phi_2$ be the angle between the rays $r_1$ and $r_2$, where $\phi_1$ and $\phi_2$ are the orientation measured for the ray $r_1$ and $r_2$, respectively (see figure 2). Assuming bounded errors for $\phi_1$ and $\phi_2$, the real value of $\theta_{12}$ is within the uncertainty interval $[\theta_{12}^-, \theta_{12}^+]$, where

$$\theta_{12}^- = \theta_{12} - \phi_{1error} - \phi_{2error}$$

$$\text{and } \theta_{12}^+ = \theta_{12} + \phi_{1error} + \phi_{2error}$$

Now, let us suppose that the ray $r_1$ matches the landmark $m_i$, making the correspondence pair $(r_1, m_i)$. Then, a landmark $m_j$ becomes a candidate to match $r_2$ only if the uncertainty ring ${}^{12}S_{ij}$ delimited by two circles $C_{ij}^+$ and $C_{ij}^-$ and defined by the parameters $(\theta_{12}^+, \theta_{12}^-, m_i, m_j)$, intersects the uncertainty region $C_0$, that is, if $C_0 \cap {}^{12}S_{ij} = R_2 \neq \varnothing$ (see figure 3a).

Similarly, given a generic ray $r_k$, the landmark $m_q$ is considered to be a consistent candidate to match $r_k$ only if:

$$R_{k-1} \cap {}^{hk}S_{pq} \neq \varnothing$$

where $R_{k-1}$ is the constraint region obtained once the first k-1 rays have been matched, and ${}^{hk}S_{pq}$ is the uncertainty ring defined by $(\theta_{hk}^+, \theta_{hk}^-, m_p, m_q)$, with $(r_h, m_p)$ being whatever correspondence pair previously established. Figure 3b illustrates this case for k=3 and h=2.

## 3. Constructing the Interpretation Tree

Taking into account the above mentioned region-based approach, the proposed algorithm constructs a tree with all the possible admissible interpretations, named the *interpretation tree*. Following the terminology used by Grimson [6], the interpretation tree is a graph where a level "i" represents all possible landmarks that could be matched with the ray $r_i$ according to the constraint region derived from the previous hypothesized matches. Thus, a complete branch, from the root to a leaf, represents an admissible interpretation for all the rays extracted from the image. In order to account for rays that do not correspond to any known landmark the algorithm includes a null landmark ("∅").

To efficiently construct the interpretation tree, the algorithm makes use of three additionally constraints: landmark ordering, prediction of visible landmarks, and attribute compatibility.

**Landmark ordering.** The map is assumed to be a single connected polygon where the landmarks are the vertices. Thus, we can create a circular list sorted on the basis of the relative connection between them. This landmark ordering permits to prune significantly the interpretation tree, since they always have to appear in the image in a given order. For example, if a ray $r_i$ is paired to a landmark $m_j$, then $r_{i+1}$ (the following ray on the right of $r_i$) must find its correspondence pair among the landmarks on the right of $m_j$. Obviously, this assumption limits the landmarks stored in the map to those connected by opaque surfaces (i.e. walls).

**Prediction of visible landmarks.** Since only a small portion of the map (in general) is visible from a particular position and orientation of the robot, landmark prediction greatly decreases the size of the interpretation tree by reducing the number of landmarks that can possibly match the extracted rays. As the absolute orientation $\eta_i$ of a given ray $r_i$ is affected by both errors in robot pose and errors in the orientation angle $\phi_i$, the candidate landmarks to match to $r_i$ are within the region of the map delimited by the lines $r_i^+$ and $r_i^-$, which are defined by the worst-case angles given by (figure 4):

$$\eta_i^+ = \alpha_o + \alpha_{error} + \phi_i + \phi_{ierror}$$

$$\eta_i^- = \alpha_o - \alpha_{error} + \phi_i - \phi_{ierror}$$

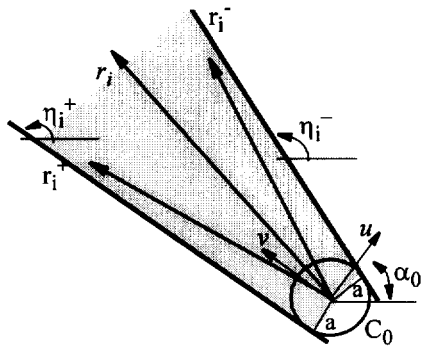



Figure 4.- Region where any candidate landmark to match the ray $r_i$ has to lie.

**Attribute compatibility**. The motivation of using attributes is not that of permitting the individual matching of each ray, but discarding landmarks that are incompatible with the attributes extracted for a given ray. In particular, we consider the following two attributes:

- *Vertical position* of the landmarks. The endpoints of vertical edge associated to a ray should be within the projection of the landmark onto the image plane.
- *Shading* in both sides of the landmarks (left and right). Three different situations may occur: "dark-to-bright", "bright-to-dark" and "undefined". The latter indicates that the landmark may appear in the image indistinctly either as a positive discontinuity (dark-to-bright) or as negative (bright-to-dark), depending on the environmental lighting conditions and on the observation angle. Of this type are, for example, the corners of the walls. This attribute is compared with the sign of the second derivative of the vertical edge associated to the ray.

Besides increasing the efficiency in the construction of the tree, the attribute compatibility provides robustness to the algorithm by weeding out inconsistent pairings.

## 4. Computing the Robot Pose

Once the interpretation tree has been constructed, all its branches are traversed in order to select the best interpretation. We suppose it is the one which contains the largest set of pairs, let say "n". From this interpretation, the robot pose $\mathbf{p} = (x, y, \alpha)$ is estimated by minimizing the error function:

$$\min_{(x,y,\alpha)}(\mathbf{e}^T\mathbf{e}) = \min_{(x,y,\alpha)}\left(\sum_{k=1}^{n} e_k^2\right) \qquad (1)$$

where $e_k = \sin(\delta_k)$ is the angular error involved in the pair $\wp_k = (r_j, m_i)$ which is given by (see figure 5):

$$\sin\delta_k = \frac{(x_i - x)\cos(\alpha + \phi_j) + (y_i - y)\sin(\alpha + \phi_j)}{\sqrt{(x_i - x)^2 + (y_i - y)^2}} \qquad (2)$$

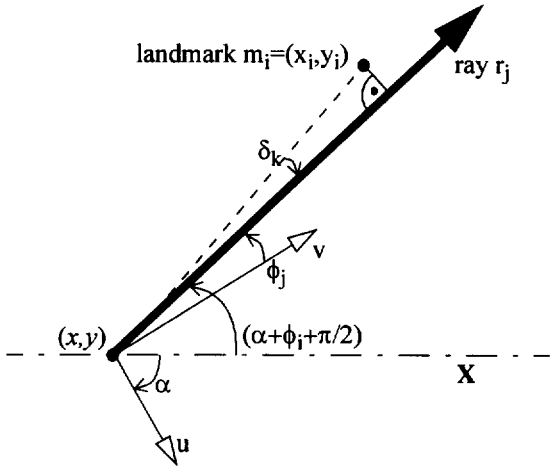



Figure 5.- Angular error $\delta_k$ involved in the pair $(r_j, m_i)$.

Since this optimization problem is non-linear we use the Gauss-Newton iterative algorithm to solve it [7]. In this method, solving (1) is equivalent to find a solution of

$$\mathbf{e} + \mathbf{J}\Delta = \mathbf{0} \qquad (3)$$

where $\mathbf{e}$ is the error vector, $\Delta$ is the difference vector between the transformation parameters on successive iterations ($p_t = p_{t-1} + \Delta_{t-1}$), and $\mathbf{J}$ is the Jacobian:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial e_1}{\partial x} & \frac{\partial e_1}{\partial y} & \frac{\partial e_1}{\partial \alpha} \\ \cdots & \cdots & \cdots \\ \frac{\partial e_n}{\partial x} & \frac{\partial e_n}{\partial y} & \frac{\partial e_n}{\partial \alpha} \end{bmatrix}$$

obtained through the partial derivatives of equation (2). In order to the algorithm to converge, the initial pose $p_0$ must lie in the uncertainty region, for example a vertex of the solution region G.

Since equation (3) is overdetermined for n>3 we use the pseudoinverse of the Jacobian to find a least square fit of $\Delta$:

$$\Delta = -(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T\mathbf{e} \qquad (4)$$

Equation (4) is solved iteratively for the displacement vector $\Delta$ until the absolute value of its elements is less than some tolerance or the number of iterations exceeds a preset value.

In the event that two or more interpretations reach the same number of pairs, all of them are considered as candidate and the one that provides the minimum of the cost function is selected.

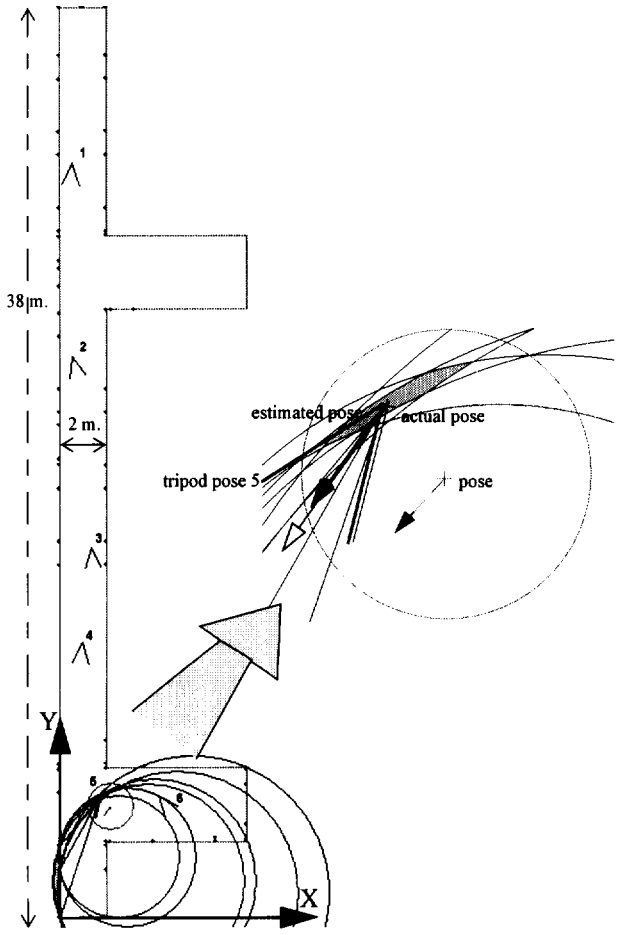



Figure 6.- Map of the hallway used in the experiments where the six poses of the tripod are marked. The lower part of the figure illustrates the circles corresponding to the best interpretation of the rays obtained at pose 5 (see figure 2).

## 5. Experimental Results and Discussions

The algorithms developed in this research have been tested in the scenario shown in figure 7, whose map containing 64 landmark appears in figure 6. The vision system consists of a CCD-camera with a 8 mm. objective connected to a frame-grabber with 768 x 576 pixels of resolution. For precision purposes and to facilitate the survey, the camera has been mounted on a calibrated tripod, instead of our mobile robot. So that the tripod poses were easily surveyed using a tape measure and a standard protractor.

The odometric poses were chosen as the surveyed poses plus an arbitrary offset greater than 0.5m. in position and 7deg in orientation. The radius of the dead-reckoning uncertainty circle was considered to be 1m., while the orientation error was ±10deg. To account for errors in the landmarks, the errors of the extracted rays was inflated up to ±0.34 deg. (equivalent to ±6 pixels of error).

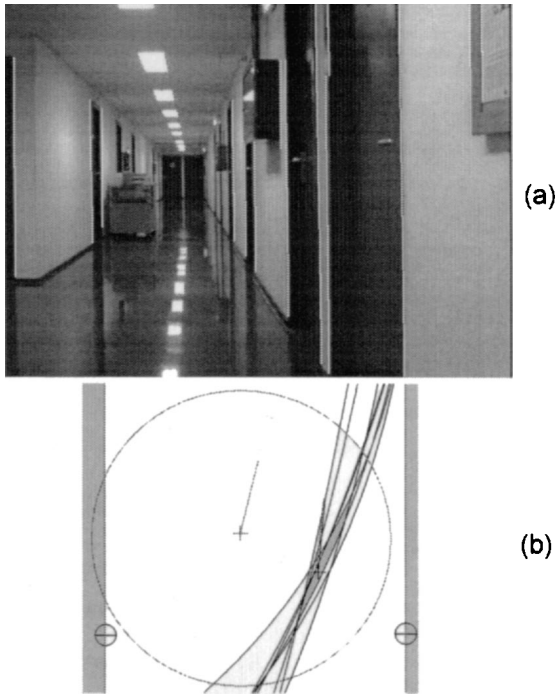



Figure 7.- (a) Image taken at location 1 showing the vertical edges extracted. The reflective vertical lines on the floor have not been detected since a significantly high threshold has been used for the line extractor algorithm. (b) The resulting uncertainty region and estimated pose.

In order to demonstrate the performance of the algorithm we have tested it in two different cases: following a path along the hallway (with many matches and long distances) and rotating about a fixed position (with few matches and short distances).

| Errors / Pose | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| X (cm.) | 1.0 | 0.1 | -0.7 | 1.1 | 2.9 | 2.7 |
| Y (cm.) | 0.5 | -0.7 | -1.1 | -3.2 | -4.5 | 1.8 |
| α (deg.) | -0.2 | -0.08 | -0.2 | -0.45 | -0.78 | -0.9 |

Table 1: Errors in the 6 estimated poses along the hallway.

In the first experiment, the tripod was manually placed at the six poses marked in the map of figure 6. The errors in the estimated poses, shown in table 1, are extremely small for the first four locations since a significant number of matches, almost symmetrically distributed along the hallway, are established[2]. Also, this symmetry gives rise to uncertainty regions appreciably aligned with the corridor

as shown in figure 7a.

| Errors / α | 0° | 30° | 150° | 160° | 170° | -160° | -110° | -90° |
|---|---|---|---|---|---|---|---|---|
| X (cm.) | -2.4 | -5.6 | 2.9 | 6.4 | 1.2 | 1.2 | -1.9 | 0.6 |
| Y (cm) | -5.7 | 3.3. | -4.5. | 1.1 | -3.4 | 3.5 | 7.5 | -1.1 |
| α (deg.) | -0.5 | -1.2 | -0.7 | -1.1 | -0.5 | -0.7 | -1.0 | -0.2 |

Table 2: Errors in the 8 estimated poses at the location 5 of the path (see text).

In the second experiment the tripod was situated at the location number 5 of the path, where eight images were taken for different orientations. As shown in the table 2, the estimated errors were worse than in the first experiment. It makes sense because these locations as well as the poses 5 and 6 of the experiment 1 are characterized by landmarks at relative short distances from the camera. Since the angular errors in determining the rays ($\phi_{error}$) decrease with the distance from the landmark to the camera[3], these poses become rather vulnerable to these errors. Another fact that influences the quality of the resulting estimates is the number of matches established, that in these cases have been appreciably smaller. This leads to larger solution regions as the one shown in figure 8.

| Pose | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| # ray | 11 | 14 | 9 | 16 | 5 | 11 |
| # landmarks checked | 52 | 38 | 25 | 24 | 10 | 7 |
| # matches | 10 | 10 | 9 | 15 | 3 | 4 |
| # false positive | 1 | 5 | 0 | 1 | 2 | 7 |
| runtimes (sec.) | 0.25 | 0.03 | 0.01 | 0.33 | 0.01 | 0.01 |

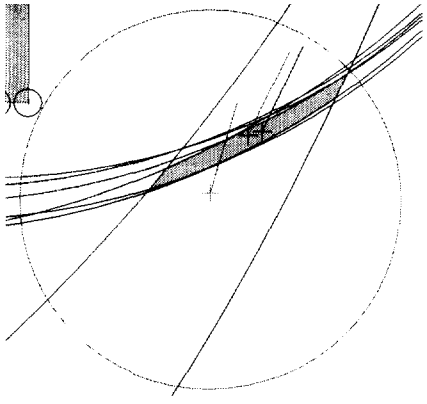
Table 3: Runtimes of the algorithm on a pentium 166 for the 6 poses along the hallway.

The computational cost grows exponentially with the number of rays, however the algorithm runs very quickly (see table 3). It is important to note that despite a significant number of both non-detected landmarks and false positives were present in the images, the algorithm came up with the correct interpretation in all the cases. In addition, as expected, the surveyed poses were inside the uncertainty region.

2.- Be aware that most of these errors are within the range of the survey procedure.

3.- This fact has also been observed by Krotkov [3].

(a)



(b)

Figure 8.- (a) Image taken at location 5 with -110 deg. of orientation, showing the vertical edges extracted. (b) The resulting uncertainty region and estimated pose.

## References


1. Talluri, R. and J.K. Aggarwal."Position Estimation Techniques for an Autonomous Mobile Robot- A review", in *The Handbook of Pattern Recognition and Computer Vision,* edited by C.H. Chen *et. al. Singapore: World Scientific*, PP. 769-801. (1993).
2. Sugihara, K. "Some Location Problems for Robot Navigation using a Single Camera", *Comput. Vision, Graphics, Image Process.*, vol. 42, no. 1, pp. 112-129, Apr. (1988).
3. Krotkov, E. "Mobile Robot Location using a Single Image". *IEEE Int. Conf. on Robotics and Automation.* pp.978-983, (1989).
4. Sutherland, K. and W. Thompson. "Localizing in Unstructured Environments: Dealing with the Errors". *IEEE Transactions on Robotics and Automation*, vol. 10, no. 6, pp 740-754, Dec. (1994).
5. Atiya, S. and G.D. Hager. Real-Time Vision-based Robot Location. *IEEE Transactions on Robotics and Automation*, vol. 9, no. 6, pp 785-800, Dec. (1993)
6. Grimson, W.E.L."Object Recognition by Computer: The Role of Geometric Constraints". *The MIT Press.* (1990).
7. Gelb, A. "Applied Optimal Estimation". *The MIT Press.* (1974).