Proceedings of the 2003 IEEE
International Conference on Robotics & Automation
Taipei, Taiwan, September 14-19, 2003

Optimal Landmark Configuration for Vision-Based Control of Mobile Robots

Darius Burschka, Jeremy Geiman, and Gregory Hager
Computational Interaction and Robotics Laboratory
Johns Hopkins University
Baltimore, MD 21218
{burschka|jag|hager}@cs.jhu.edu

***Abstract*— We analyze the problem of finding the optimal placement of tracked primitives for robust vision-based control of a mobile robot. The analysis evaluates the properties of the Image Jacobian matrix, used for direct generation of the control signals from the error signal in the image, and the accuracy of the underlying sensor system. The analysis is then used to select optimal tracking primitives that ensure good observability and controllability of the mobile system for a variety of sensor system configurations.**

The theoretical results are validated with our mobile robot for system configurations that use standard video cameras mounted on a pan-tilt head and catadioptric systems.

I. MOTIVATION

Localization is a fundamental task on mobile systems. Knowledge about the current position relative to landmarks in a local area is essential for planning and task specification. In many cases it is not necessary to use sophisticated models of the environment to specify the path of a mobile system. Many systems merely need to follow pre-specified paths learned in a *teaching* phase to fulfill their tasks, e.g. mail delivery robots, storage management robots, sentry robots, etc. This is in fact our goal: to develop a system that can be *walked* in a teaching phase through the environment. During this phase it learns the path based on position of identified tracking primitives, like color blobs, gray-scale patterns or disparity regions. It uses this knowledge later to repeat this path by generating the control signal directly from the error between the expected position in the image for the tracked primitives and their actual position during the *replay* step.

A standard localization method is to measure angles between landmark positions and to compute the pose from intersections of the circles that represent the possible positions from each possible pair of bearing measurements [14], [10], [6], [2]. We extend this idea by measuring both azimuth and elevation angles of a landmark to represent its position on a sphere instead of a circle (Fig. 1).

There have been a number of papers on the process of selecting useful features points or natural markers in image data [7], [8], [9], [11], [12], [13], [15]. These approaches select optimal landmarks based on their appearance in the image. In this paper we analyze the problem of finding the optimal placement of tracked landmarks based on their 3D position in the world.

In [3] we presented a system that allows vision-based control of mobile robots in a local path segment with pre-selected tracking primitives (*features*). The system directly uses the input from a sensor that provides measurements of the angular directions of the incident light rays to navigate in a local area (Fig. 1). This ideal sensor can be approximated with a variety of sensor configurations such as omnidirectional cameras or standard cameras mounted on pan-tilt heads, to give just a short list of possible configurations.

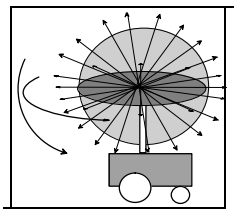

Fig. 1. Ideal generic sensor system measuring the horizontal α and vertical β angle of incidence.

Theoretical considerations presented in this paper are validated in multiple experiments. They show that varying landmark configurations do significantly influence the quality of the generated control signals. Finding optimal landmark configurations is essential for several areas of work with mobile systems. The correct selection depends partly on spatial resolution of the used sensor system. The results from this paper help to choose an optimal sensor system for a specific room structure with pre-specified artificial landmarks. In on-line landmark selection process during the initial phase of operation in a local path segment, the results of this paper help to choose landmarks allowing best control signals in a given segment.

Our goal is to investigate the properties of the Jacobian matrix used to map deviations in the 3D space onto changes in the perception of the ideal sensor (Fig. 1).

The remainder of this article is structured as follows. The next section describes the geometry of the vision-based control problem. Section III describes the evaluation

of the system sensitivity to changes in the input parameters and the expected input sensitivities of the camera systems. Section IV presents numerical evaluation and experimental results from a real system. We close with a discussion of future work.

II. Vision-Based Control System

As already described in [3] in more detail, the presented navigation system operates in two steps. In a *teaching* phase the user takes the robot for a *walk*. The robot saves the image positions of selected tracking primitives (e.g. color blobs, gray-scale patterns, etc.) to identify positions in the world that later are used to repeat this path autonomously in the *replay* phase. The control signals are generated directly from the error signal in the image using the Image Jacobian matrix described in Section II-B.

A. Spherical Image Projection

We assume a non-holonomic mobile system with unicycle kinematics throughout the article. The system operates in the (x, z, Θ) coordinates (Fig. 2).

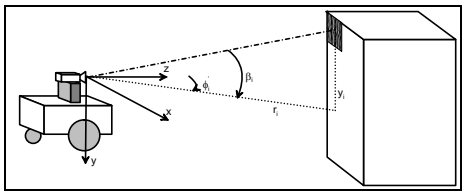

Fig. 2. Coordinate system used in the system.

We describe the imaging properties of the generic sensor system depicted in Fig. 1 in spherical coordinates (α - azimuth angle, β - elevation angle). The origin of the robot coordinate system is assumed to be at the center of rotation of the mobile system, the optical z axis points in the "forward" direction of the robot motion, and the x axis points to the "right" of the robot orientation (Fig. 2). A point in space relative to the robot can then be described by the triple (x_i, y_i, z_i).

We define the spherical coordinates (α_i, β_i) in the camera projection of an observed point P_i to

$$\alpha_i = \arctan\frac{x_i}{z_i} \;\wedge\; \beta_i = \arctan\frac{y_i}{\sqrt{x_i^2 + z_i^2}} \qquad (1)$$

B. The Image Jacobian

Now, assuming holonomic motion in the plane, we can compute the following Image Jacobian that relates the change of angles in the image, (α_i, β_i), to changes in position in the (x, z)-plane from (1):

$$\mathcal{J}_i^t = \begin{pmatrix} \frac{\partial\alpha_i}{\partial x} & \frac{\partial\alpha_i}{\partial z} & \frac{\partial\alpha_i}{\partial\Theta} \\ \frac{\partial\beta_i}{\partial x} & \frac{\partial\beta_i}{\partial z} & \frac{\partial\beta_i}{\partial\Theta} \end{pmatrix} = \begin{pmatrix} \frac{z_i}{x_i^2+z_i^2} & -\frac{x_i}{x_i^2+z_i^2} & -1 \\ -\frac{x_i y_i}{(x_i^2+y_i^2+z_i^2)\cdot\sqrt{x_i^2+z_i^2}} & -\frac{y_i z_i}{(x_i^2+y_i^2+z_i^2)\cdot\sqrt{x_i^2+z_i^2}} & 0 \end{pmatrix} \qquad (2)$$

The dependency on the unknown position (x_i, z_i) of the robot relative to the tracked landmark can be avoided considering the geometry of the system to:

$$\begin{aligned} &\sqrt{x_i^2 + y_i^2 + z_i^2} = \tfrac{y_i}{\sin\beta_i},\ \sqrt{x_i^2 + z_i^2} = \tfrac{y_i}{\tan\beta_i}, \\ &x_i = \tfrac{y_i\cdot\sin\alpha_i}{\tan\beta_i}, \qquad z_i = \tfrac{y_i\cdot\cos\alpha_i}{\tan\beta_i} \\ &\mathcal{J}_i^t = \begin{pmatrix} \frac{\tan\beta_i\cdot\cos\alpha_i}{y_i} & -\frac{\tan\beta_i\cdot\sin\alpha_i}{y_i} & -1 \\ -\frac{\sin^2\beta_i\cdot\sin\alpha_i}{y_i} & -\frac{\sin^2\beta_i\cdot\cos\alpha_i}{y_i} & 0 \end{pmatrix} \end{aligned} \qquad (3)$$

Note in particular that the Image Jacobian is a function of only one unobserved parameter, y_i, the height of the observed point. Furthermore, this value is *constant* for motion in the plane. Thus, instead of estimating a time-changing quantity as is the case in most vision-based control, we only need to solve a simpler static estimation problem. We refer to [3] for a detailed description of the y_i-estimation.

In the following text we assume that the system already learned the positions of the tracked objects as columns in a matrix $\mathcal{M}_p$ in the *teaching* phase [3]. In the *replay phase* the values $\mathcal{M}_p[t][i]$ representing the stored positions for the tracker i at the time stamp t together with the estimations of y_i are used to calculate the position error of the robot $\Delta w^t = (dx, dz, d\Theta)^T$.

For each tracked landmark we can write the dependency of the observation error in the image Δe_i^t on the position error Δw^t using (2) to

$$\Delta e_i^t = \mathcal{J}_i^t \cdot \Delta w^t, \quad \text{with} \quad \Delta e_i^t = \mathcal{M}_p[t][i] - p_i^t. \qquad (4)$$

Since all observations depend on the same position error Δw^t and we are interested in estimation of the position error Δw^t from the error in the camera image Δe_i^t, we need to invert the equation (4). It is not possible to estimate all three values of Δw^t from one landmark (α_i, β_i) in the image. We compute a "stacked" observation vector Δe^t using (4) to

$$\begin{aligned} \Delta e^t &= \mathcal{J}^t \cdot \Delta w^t, \quad \text{with} \\ \mathcal{J}^t &= (\mathcal{J}_1^t, \ldots, \mathcal{J}_N^t)^T \;\wedge\; \Delta e^t = (\Delta e_1^t, \ldots, \Delta e_N^t)^T \end{aligned} \qquad (5)$$

From (5) we can estimate Δw^t using the pseudo-inverse $(\mathcal{J}^t)^{-1}$ of the "stacked" Image Jacobian matrix from (2) to

$$\Delta w^t = (\mathcal{J}^t)^{-1} \cdot \Delta e^t, \quad \text{with} \quad (\mathcal{J}^t)^{-1} = (\mathcal{J}^{t^T}\mathcal{J}^t)^{-1}\mathcal{J}^{t^T} \tag{6}$$

The value Δw^t describes the error in the 3D position that we use to generate the control signals for the robot.

III. System Analysis

The relative error in the solution caused by perturbations of parameters can be estimated from the condition number of the Image Jacobian matrix J. The condition number is the ratio between the largest and the smallest singular value of the matrix J.

The condition number estimates the sensitivity of solution of a linear algebraic system to variations of parameters in matrix J and in the measurement vector b.

Consider the equation system with perturbations in matrix J and vector b:

$$(\mathcal{J} + \epsilon\delta\mathcal{J})x_b = b + \epsilon\delta b \tag{7}$$

The relative error in the solution caused by perturbations of parameters can be estimated by the following inequality using the condition number κ calculated for J (see [5]):

$$\frac{||x - x_b||}{||x||} \leq \kappa \left(\epsilon \frac{||\delta\mathcal{J}||}{||\mathcal{J}||} + \epsilon \frac{||\delta b||}{||b||} \right) + \mathcal{O}(\epsilon^2) \tag{8}$$

Therefore, the relative error in solution x can be as large as condition number times the relative error in J and b. The condition number together with the singular values of the matrix J describe the sensitivity of the system to changes in the input parameters.

In the following subsections we investigate the observability and accuracy of the output parameters (x, y, z) from the input stream of the camera system (sec. III-A) and the influence of the real sensor on the achievable accuracy of the system (sec. III-B).

A. Optimal Landmark Configuration for the Image Jacobian Matrix

The singular values can be obtained as positive square roots of the eigenvalues of the matrix $J^T \cdot J$. With $y_{i\in\{1,\ldots,N\}}$ as heights of the tracked objects, $\alpha_{i\in\{1,\ldots,N\}}$ as azimuth angles to them and $\beta_{i\in\{1,\ldots,N\}}$ as their elevation angles. The resulting matrix for N landmarks has the form shown in (9).

The system estimates three parameters $(dx, dy, d\Theta)$ from the image positions (u_i, v_i) of all tracked primitives (features) $i \in \{1, \ldots, N\}$ (4). Therefore, at least two *features* are necessary to estimate all 3 position parameters.

Each feature contributes a measurement of a distance Δr_i from the robot to the feature in the ground plane and an orientation $\Delta\Theta$ relative to it. The equation (3) can then be written in this case as:

$$\mathcal{J}_i^t = \begin{pmatrix} 0 & -1 \\ -\frac{\sin^2\beta_i}{y_i} & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -\frac{1}{y_i \cdot \left(1 + \left(\frac{r_i}{y_i}\right)^2\right)} & 0 \end{pmatrix}$$

$$r_i = \sqrt{x_i^2 + z_i^2} \tag{10}$$

From the equation (10) we learn that an error $\Delta\Theta$ is directly forwarded to the output value α_i, while the value Δr, the error in the distance to the feature, is scaled with the value

$$\kappa_r = \left[y \cdot \left(1 + \left(\frac{r_i}{y_i} \right)^2 \right) \right]^{-1} \tag{11}$$

Since in our case the measurement error is in the image space, the resulting errors in the world are dependent on the reciprocal values.

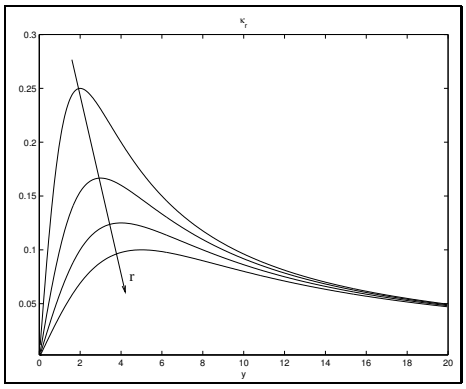

Fig. 3. Dependency of κ_r on y_i and r_i.

We deduce from the above equation that the optimum placement of the feature should maximize the above expression to allow good observability of the position error Δr. The optimal value can be estimated to

$$\begin{aligned} \frac{d\kappa_r}{dy_i} &= -\frac{1}{{x_i}^2\left(1 + \frac{{r_i}^2}{{y_i}^2}\right)} + \frac{2 \cdot {r_i}^2}{{y_i}^4\left(1 + \frac{{r_i}^2}{{y_i}^2}\right)^2} = 0 \\ \Rightarrow \quad y_i &= \pm r_i \quad \Rightarrow \beta_i = \arctan\frac{y_i}{r_i} \end{aligned} \tag{12}$$

that corresponds to an angle $|\beta_i| = 45°$.

It is shown that any linear system has at least one solution component whose sensitivity to perturbations is proportional to the condition number of the matrix, but there may exist many components that are much better conditioned [4]. The sensitivity for different components of the solution changes depending on the configuration of the landmarks and the relative position of the robot to them.

$$\mathcal{J}^T \cdot \mathcal{J} = \begin{pmatrix} \sum_{i=1}^{N} \left(\frac{\tan^2 \beta_i \cdot \cos^2 \alpha_i}{y_i^2} + \frac{\sin^4 \beta_i \cdot \sin^2 \alpha_i}{y_i^2} \right) & \sum_{i=1}^{N} \left(\frac{\sin^4 \beta_i \cdot \sin \alpha_i \cdot \cos \alpha_i}{y_i^2} - \frac{\tan^2 \beta_i \cdot \cos \alpha_i}{y_i^2} \right) & \sum_{i=1}^{N} \left(-\frac{\tan \beta_i \cdot \cos \alpha_i}{y_i} \right) \\ \sum_{i=1}^{N} \left(\frac{\sin^4 \beta_i \cdot \sin \alpha_i \cdot \cos \alpha_i}{y_i^2} - \frac{\tan^2 \beta_i \cdot \cos \alpha_i}{y_i^2} i \right) & \sum_{i=1}^{N} \left(\frac{\tan^2 \beta_i \cdot \sin^2 \alpha_i}{y_i^2} + \frac{\sin^4 \beta_i \cdot \cos^2 \alpha_i}{y_i^2} \right) & \sum_{i=1}^{N} \left(\frac{\tan \beta_i \cdot \sin \alpha_i}{y_i} \right) \\ \sum_{i=1}^{N} \left(-\frac{\tan \beta_i \cdot \cos \alpha_i}{y_i} \right) & \sum_{i=1}^{N} \left(\frac{\tan \beta_i \cdot \sin \alpha_i}{y_i} \right) & N \end{pmatrix} \tag{9}$$

Figure 4 shows that the sensitivity to perturbations in image coordinates is highest when close to the two landmarks selected for this evaluation. The two peaks in Figure 4 indicate this. The sensitivity to changes in the x-coordinate are slightly better (left) than the sensitivity to the changes in z direction (right in Figure 4).

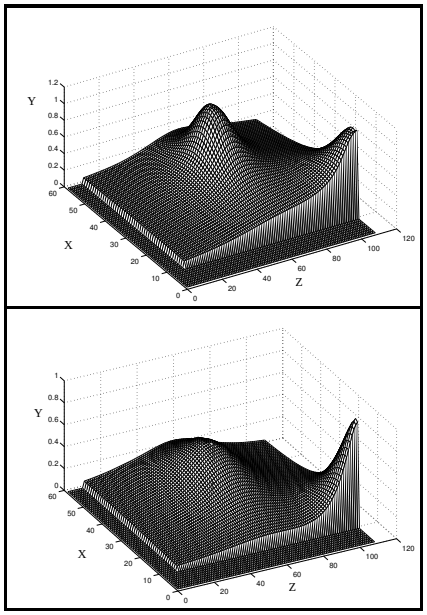

Fig. 4. Sensitivity (Y) to changes in x (top) and z (bottom) coordinates for two landmarks at (3,2,5) and (0,1,10).

This result could be validated in the result section with a real sensor system (section IV-B).

Figure 5 shows the expected uniform sensitivity to angular errors $d\Theta$ (10). The visible singularities are at the positions of the tracked landmarks, where a simultaneous measurement of the horizontal components of both landmarks is not possible.

B. Angular Resolution of the Sensor System

We assumed so far in this paper that the used sensor system meets the requirement from Figure 1 of uniform angular resolution in the α and β directions.

This ideal sensor can only be approximated with real, physical sensor systems. In our experiments we used two different sensor configurations that approximate this ideal sensor (Fig. 6).

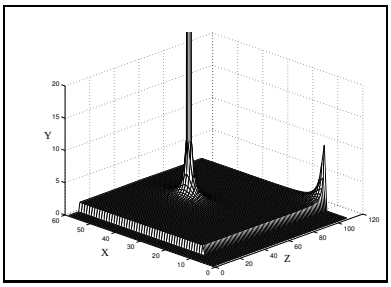

Fig. 5. Sensitivity (Y) to changes in Θ.

Fig. 6. Two implementations of the ideal sensor from Figure 1: (left) combination of a pan-tilt head with standard camera; (right) omnidirectional camera.

The standard video camera has a limited field of view. A camera equipped with an 8mm lens has a field of view of approximately $50°$. The angular resolution of this camera varies depending on the distance from the optical center and is highest around the center. The resolution of this camera ϵ_p around center can be estimated to

$$\epsilon_p = \arctan \frac{p_x}{f}, \; \mathrm{p_x} - \mathrm{pixelsize}, \; \mathrm{f} - \mathrm{focal\,length} \tag{13}$$

This type of camera is very sensitive to coordinate changes in the world, especially around the optical center. However, the limited field of view restricts the possible configurations of the landmarks to just a narrow field in front of the camera. Considering the requirement for the landmarks to be visible over a long path segment, this kind of camera gives poor control results at the beginning of the segment, when the landmarks are still far away. Objects at larger distances away appear closer to the center

where the resolution is higher than in the periphery. The characteristics of the angular resolution for this type of camera slightly compensates the decrease in the sensitivity of the Jacobian matrix for larger distances from the robot.

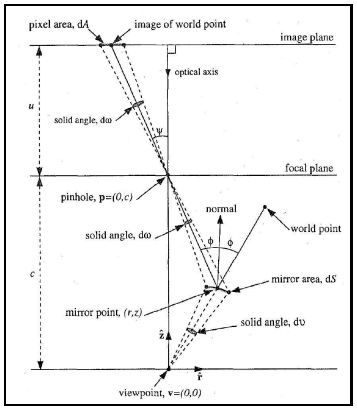

Fig. 7. The geometry used to derive the resolution of a catadioptric sensor (from [1]).

According to the derivation in [1] the resolution of a catadioptric camera (i.e. omnidirectional camera) $dA/d\nu$ with a hyperbolic mirror relates to the resolution of a standard camera $dA/d\omega$ as (see Fig. 7):

$$\frac{dA}{d\nu} = \frac{r^2 + z^2}{(c - z)^2 + r^2} \cdot \frac{dA}{d\omega} \tag{14}$$

where (r, z) is a point on the mirror being imaged and c is the distance between the viewpoint of the mirror-camera configuration and the focal point of the lens. Using simple properties of hyperboloids it follows that the factor in (14) increases with r. This system has the highest resolution around the periphery. In the configuration depicted in Figure 6 objects in the periphery are in the largest distance from the robot. Here again the imaging properties of the camera system compensate for the flaws in the sensitivity of the control system. The omnidirectional camera has the advantage to track features over a larger azimuth range of $360°$ compared to the $50°$ of a standard camera. This allows landmarks to be much closer to the robot during the initial landmark selection. Additionally, it allows them to stay visible over the same path length since the robot can pass between them without losing track of any of them. Since the sensitivity of the system decreases quadratically with the distance r_i to the landmark (11), reducing the initial distance results in a high sensitivity gain of the system observed on our mobile system.

IV. Results

A. Sensitivity of the Jacobian Matrix

Based on the considerations in Section 3 we conclude that the system tries to minimize the distances to the tracked features to keep κ_r maximal (see (11) and Fig. 3).

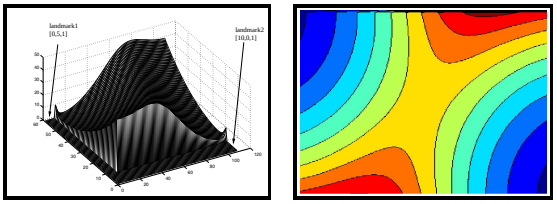

Fig. 8. Distribution of the condition number values for 2 similar landmarks:(left) surface plot; (right) contour plot.

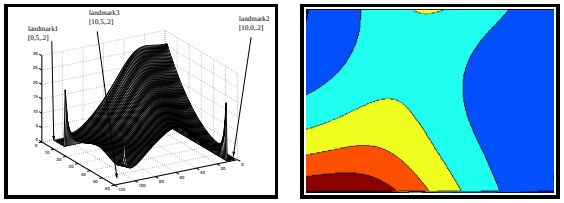

Fig. 9. Distribution of the condition number value for 3 similar landmarks: (left) surface plot; (right) contour plot.

The plotted examples for two and three landmarks in Figures 8 and 9 confirm this assumption. Each landmark results in a cavity in the plotted surface representing the condition number.

The influence of a single landmark on the condition number can be evaluated by choosing two landmarks with extreme different impacts on the shape of the surface. In the following example we have chosen two landmarks at the coordinates (3,4,5) and (0,0.1,10). We follow from equation (11) that the influence of the second landmark ($y_2 = 0.1$) is several magnitudes lower than that of the first landmark ($y_1 = 5$). This assumption is verified in Figure 10.

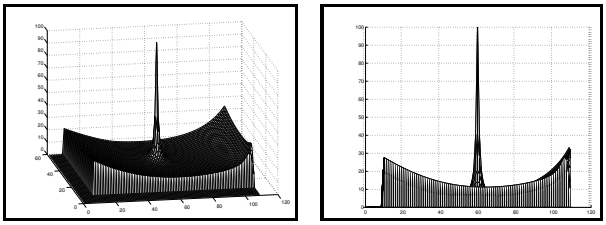

Fig. 10. Influence of a single landmark on the condition number: (left) perspective view; (right) projection from one side.

B. Overall Sensitivity of the System

As mentioned already in Section III, the error of the system output consists of the error in the system input projected by the Jacobian matrix onto the output data space. For the sensor configuration consisting of an omni-directional sensor that measures the angles of the incident rays, we measured the input error in the image space (Fig. 11).

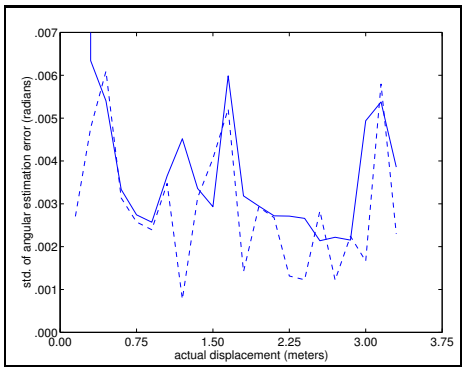

Fig. 11. Standard deviation of the error in the estimation of α (solid) and β (dashed) angles (in radians) from the tracking process vs. displacement as system moves away from tracked features.

As already predicted in Section III-A, the x-coordinate estimates better follow the changes in the image coordinates than the z-coordinate estimates (Fig. 12). The sensitivity for the x-coordinates is higher and allows for a higher accuracy of control in this direction.

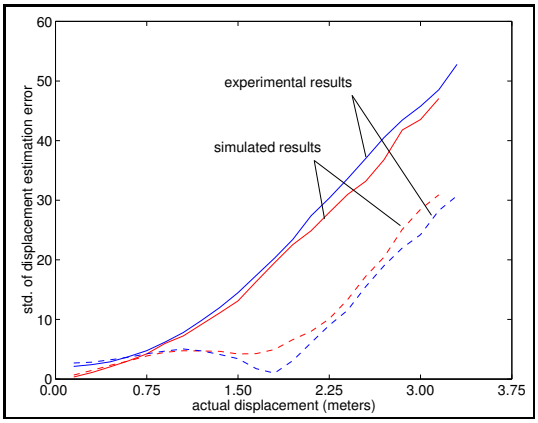

Fig. 12. Standard deviation of the error in the estimation of the displacement for x (solid) and z (dashed) as system moves away from tracked features. Experimental and simulated results shown.

Fig. 12 validates the correctness of the used system model. The predicted error for both coordinates (x, z) matches the actual measurements on our mobile system.

V. Conclusions

Our work is part of a system for sensor based navigation in which natural landmarks are used as markers or landmarks. In this system, a set of natural landmarks is used for navigation in a given path segment. The sets change each time any of the landmarks disappear due to occlusions or limited field of view of the conventional camera system. This defines the boundaries between the path segments. The landmarks can be selected manually by the user in a teaching phase or they are selected automatically by the system.

We have discussed the optimal selection of landmarks to obtain the best control results for a given environment based on the sensitivity of the sensor system used to model the ideal sensor with uniform angular resolution from Figure 1 and based on projection properties of the Image Jacobian matrix used to generate the control signals from the image data.

The presented results can be used for optimal configuration of catadioptric and standard camera systems. A correct selection of the imaging properties of the camera can compensate some of the insensitivities of the control system such as for landmarks at larger distances from the robot.

In the future, the system will be extended to automatically select new landmarks for omnidirectional navigation based on the results from this paper.

Acknowledgments

This work was supported by the MARS project, and by the NSF RHA program.

VI. References

[1] S. Baker and S.K. Nayar. A Theory of Catadioptric Image Formation. *In Proc. of IEEE International Conference on Computer Vision*, 1998.

[2] M. Betke and L. Gurvits. Mobile Robot Localization Using Landmarks. *IEEE Trans. on Robotics and Automation*, 13:251–263, 1997.

[3] D. Burschka and G. Hager. Vision-based control of mobile robots. In *Proc. International Conference on Robotics and Automation*, pages 1707–1713, 2001.

[4] S. Chandrasekaran and I. C. F. Ipsen. On the Sensitivity of Solution Components in Linear Systems of Equations. *SIAM Jounral on Matrix Analysis and Applications*, 16(1):93–112.

[5] J. Demmel. The componentwise distance to the nearest singular matrix. *SIAM Journal on Matrix Analysis and Applications*, 13:10–19, 1992.

[6] U.D. Hanebeck and G. Schmidt. Set theoretic localization of fast mobile robots using angle measurements. *IEEE Conf. on Robotics and Automation*, page 1387, 1996.

[7] S. Se, D. Lowe, and J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *IEEE Conf. on Robotics and Automation*, pages 2051–2058, 2001.

[8] R. Sim and G. Dudek. Mobile robot localization from learned landmarks,. In *IROS*, 1998.

[9] Saul Simhon and Gregory Dudek. Selecting targets for local reference frames. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 2840–2845, 1998.

[10] K.T. Sutherland and W.B. Thompson. Localizing in unstructured environments: Dealing with the errors. *IEEE Transactions on Robotics and Automation*, 10:740–754, 1994.

[11] Y. Takeuchi, P. Gros, M. Hebert, and K. Ikeuchi. Visual learning for landmark recognition. In *Proc. Image Understanding Workshop*, pages 1467–1473, 1997.

[12] Sebastian Thrun. Finding landmarks for mobile robot navigation. In *IEEE Conf. on Robotics and Automation*, pages 958–963, 1998.

[13] C. Tomasi and J. Shi. Good features to track. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 593–600, 1994.

[14] U. Wiklund, U. Anderson, and K. Hyypa. Agv navigation by angle measurements. *Proc. 6th Int. Conf. AGV Systems*, pages 199–212, 1988.

[15] E. Yeh and D. Kriegman. Toward selecting and recognizing natural landmarks. In *IEEE Int. Workshop on Intelligent Robots and Systems*, pages 47–53, 1995.