Technical Reports (CIS)   Department of Computer & Information Science

February 1987

Information Maps for Active Sensor Control

Greg Hager
University of Pennsylvania

Information Maps for Active Sensor Control

Abstract

This paper outlines our current progress in active sensor control. We consider the problem of controlling a nonlinear observation system observing data implicitly related to parameters of interest. We show how linear estimation theory can be applied to this problem, and develop the notion of an *information map* showing the information expected from sensor viewpoints. We discuss the robustness of these techniques, and propose a method to enhance their robustness. We expect these maps to be useful in active sensor control.

Comments

INFORMATION MAPS FOR
ACTIVE SENSOR CONTROL

Greg Hager
MS-CIS-87-07
GRASP LAB 91

Department Of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

February 1987

Information Maps for Active Sensor Control

Greg Hager*

February 9, 1987

Abstract

This paper outlines our current progress in active sensor control. We consider the problem of controlling a nonlinear observation system observing data implicitly related to parameters of interest. We show how linear estimation theory can be applied to this problem, and develop the notion of an *information map* showing the information expected from sensor viewpoints. We discuss the robustness of these techniques, and propose a method to enhance their robustness. We expect these maps to be useful in active sensor control.

1 Introduction

If robots are to perform tasks in unconstrained environments, they will have to rely on sensor information to make decisions. However, sensors are not perfect imaging devices. Signal noise and discretization effects lead to information that has some uncertainty associated with it. In an unconstrained environment, robot decision making will depend on world models built from sensor information, so the information needed for proper action will be uncertain. Thus, in order to make informed decisions, the robot will need to take action explicitly devoted to reducing uncertainty. For this to be possible, sensory systems must be controllable, or *active* [1].

Active perception frees a system from the restriction of a single, static image by assuming the sensor is free to probe and explore the environment in search of data. This removes the restriction of the single image and replaces it with the much more flexible approach of actively seeking and using several scenes or samples. Active perception does not necessarily imply that the sensors physically move in space, but rather that they have controllable parameters that are changed in an intelligent fashion to influence the data gathering process. For example,

consider a camera fixed to the end of a robot manipulator. The camera can be located in space by moving the arm, and image quality is influenced by the setting of aperture and focus.

The problem we are considering is how to control active sensors in order to reduce uncertainty about the environment. In [2], we outlined a general approach for the reduction of quantitative and qualitative uncertainty. For the purposes of this paper, uncertainty can be attributed to two causes: uncertainty due to signal noise, and incompleteness of information due to limitations of sensor scope. At the micro (signal) level, the simplicity of data and the nature of the noise leads to consideration of statistical methods of modeling and analysis of signal noise. Since any information derived from a noisy signal will itself be subject to that noise, all we will assume that all information in a model can be represented via probability distributions.

From a geometric standpoint, scope limitations provide the fundamental constraint in sensor control. The choice of viewpoint can substantially influence both the quantity and quality of information a sensor furnishes. For example, a camera must be pointed at an object to observe it. Different points of view give different information about the object. Moreover, we can expect that the quality of information varies with viewing distance, aperture and other controllable parameters.

Our approach to this problem is based on mapping the information expected from different vantage points. At this point in time, we are considering the problem of refining information. We assume that we have some prior information on an object, and we seek to control the sensor in order to improve this estimate. This paper outlines our methods for predicting the information content of camera views. We are currently developing methods for sensor control based on these "information maps."

1.1 Formalizing the Problem

A sensor can be thought of as a controllable measurement system. The choices of control parameters determines the information returned by the sensor. We can formalize a controllable measurement device as a mathematical system of the following general form:

$$z = H(u, p) + V(u) \tag{1}$$

where u is the control vector for the measurement system, and p is the quantity we are attempting to observe. We observe z, a function of both u and p contaminated with additive noise $V(\cdot)$. In general, we assume that V also varies according to our choice of control.

Our problem is to maximize, by choice of u, some measure of the information z carries about p. In control theory, the ability to estimate p based on z is referred to as the *observability* of a system. For the purpose of this paper, we will assume the observed system is static so that we are left with a point estimation problem. Then, the observability of p is basically the the *information matrix* of p resulting from the observation system.

The information matrix of a distribution is closely related to the variance covariance matrix of the distribution. Recall the variance of a univariate density f is defined as

$$\sigma^2 = \int_X (x - \bar{x})^2 f(x) dx = E[(x - \bar{x})^2]$$

In the multivariate case, we consider the variance-covariance matrix given by $E[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T]$. In the case of the Gaussian distribution, the information matrix is the inverse of the variance-covariance matrix.

1.2 Previous Work

There has been some relevant work in the control theory literature related to controllable measurement subsystems. Meier [3] deals with specializations of Equation 1 in which in which $V(\cdot)$ is constant zero-mean Gaussian noise, and the system is linear in x as in

$$z = H(u)x + V$$

In this case, the best measurement control for a dynamic system can be derived. It is shown that the optimal control is open-loop and the solution is given by a dynamic program.

Müller and Weber [4] consider the related problem of maximizing the observability/controllability of a system linear in both state and control. Their approach is to maximizing a suitable norm of the observability matrix by adjusting certain parameters of the system. The norms they discuss are the trace, determinant, an maximum eigenvalue of the observability matrix.

Our problem is distinguished by a number of important criteria. First, we may not be able to directly observe the quantity we wish to estimate. Our observations may only bear an implicit relation to the parameters of interest. Second, our systems tend to be nonlinear in both state and control. Third, our measurement noise depends on the control of the measurement system. Finally, and perhaps most importantly, our information is limited by sensor scope. Our information may vary widely and discontinuously based on what information is presented to the sensory system. Hence, we have taken the approach of first investigating the properties of typical measurement systems. Once we have an understanding of the relationship between control and information, we can solve the problem of control.

2 Statistical Information

The information matrix resulting from a number of measurements depends on the estimation technique employed. Possibly the simplest technique is the *Minimum Square Error* solution. In this case, nothing needs to be assumed other than the data has additive noise. The problem can be stated as finding a δ which minimizes:

$$\min_{\delta(\cdot)} \|\delta(z) - \theta\|^2$$

Maximum likelihood is applicable if we known a distribution, $f(\cdot)$, for the observed data. In this case, we attempt to pick $\delta(\cdot)$ so that the *probability* of the observed data is maximized. This method can be stated generally as:

$$\delta(z) = \arg\max_{\theta} f_{\theta}(z)$$

When the data has a gaussian distribution, mean square error is equivalent to maximum likelihood techniques.

If both a prior on θ and a distribution on the observations, z, are known then we can attempt to minimize the *mean square error*. That is, we attempt to find the $\delta(\cdot)$ such that:

$$\min_{\delta(\cdot)} E\|\delta(z) - \theta\|^2 \tag{2}$$

In the general case, we know that the δ that minimizes Equation 2 is

$$\delta(z) = E[\theta|z]$$

This solution looks deceptively simple, but it depends on having the joint distribution of θ and z which may be extremely complex to derive. However, if z and θ are jointly gaussian, then the problem is much simpler as $\delta(\cdot)$ is known to be an affine transformation on z. In this case, if the prior on θ is $N(\hat{\theta}, \Lambda_{\theta})$ and z has distribution $N(\mu, \Lambda_z)$, we have:

$$
\begin{aligned}
E[\theta|z] &= (I - KH)\hat{\theta} + Kz, \qquad K = \Lambda_{\theta} H^T (H \Lambda_{\theta} H^T + \Lambda_z)^{-1} & (3) \\
\Lambda_{\theta|z} &= (I - KH)\Lambda_{\theta}(I - KH)^T + K\Lambda_z K^T & (4)
\end{aligned}
$$

The general solution for linearly time-varying θ is what is referred to as the *Kalman filter*.

2.1   Deriving an Observation System Based on Constraints

In general, we may not be able to directly observe the information of interest. Instead, we are forced to make observations which bear some implicit relationship to the parameters we seek. This can be generally expressed in terms of a constraint of the form:

$$g(x, p) = 0 \tag{5}$$

4

We assume the parameter p is the desired information. Observations of x serve to reduce the degrees of freedom in p. Mathematically, p belongs to a set of points forming a *manifold*. As observations of x are taken, this set is restricted until (ideally) a single point is left. Realistically, the observations of x will be contaminated with noise, forcing us to estimate p. Thus, we would like to apply the statistical methods outlined in the previous section.

The problem is one in constrained nonlinear optimization. In general, these solutions are not analytically or computationally tractable. We must then search for a way to simplify the problem. One way of simplifying the problem is to reduce the nonlinear constraints to linear ones, and then apply the solution for linear observation systems given in Equations 3 and 4.

3   First Order Techniques

One way to avoid the complications of nonlinear optimization is to approximate the nonlinear constraint with a linear one, and then apply linear techniques. For our general constraint equation, this is done by approximating equation 5 with the first order terms of a Taylor series centered about prior estimates for x and p.

We will assume that our observations of x are taken from a system of the form

$$z = x(u) + W(u) \tag{6}$$

where $x(\cdot)$ is the true parameter, and $W(\cdot)$ represents zero-mean gaussian noise with variance-covariance Λ_W. We will further assume that p has is distributed about a prior mean $\hat{p}$ with variance-covariance Λ_p. Expanding Equation 5, we get

$$g(x,p) \approx g(z,\hat{p}) + (x-z) \left(\left. \left. \frac{\partial g(x,p)}{\partial x} \right|_{(x=z)} \right|_{(p=\hat{p})} + (p-\hat{p}) \left. \left. \frac{\partial^2 g(x,p)}{\partial p \partial x} \right|_{(x=z)} \right|_{(p=\hat{p})} \right)$$
$$+ (p-\hat{p}) \left. \frac{\partial g(z,p)}{\partial p} \right|_{(p=\hat{p})}$$

By dropping all terms of higher order than linear, this equation can be rewritten to form a linear observation system:

$$y = Mp + V, \qquad M = \left. \frac{\partial g}{\partial p} \right|_{\hat{p}} \qquad V = -\left. \frac{\partial g}{\partial x} \right|_{\hat{p}} (x-z) \tag{7}$$

Note that from Equation 6 we know $E[z-x] = 0$ and $E[(x-z)(x-z)^T] = \Lambda_W$.

$$\Lambda_V = \left. \frac{\partial g}{\partial x} \right|_{\hat{p}} \Lambda_W \left(\left. \frac{\partial g}{\partial x} \right|_{\hat{p}} \right)^T \tag{8}$$
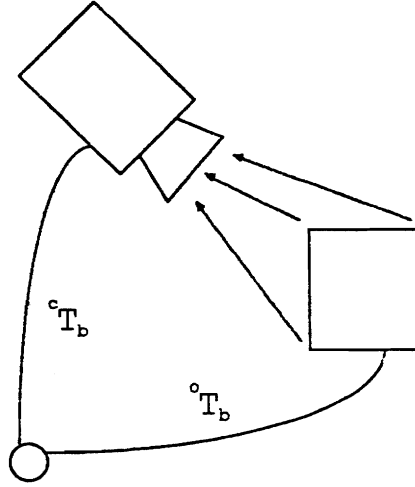
5

Figure 1: The Positioning Problem

Applying linear estimation techniques to this formulation will result in information matrices of the form:

$$\Lambda_p^{-1} = M^T \Lambda_V^{-1} M \tag{9}$$

3.1 Geometric Information Maps

We will apply the linearization methods to the problem of determining general position of an object in two dimensions. The problem can be posed as follows. Assume an observation system and an object are related to a fixed base coordinate system by [5]

$$^cT_b = \text{translate}(x_c, y_c, 0)\text{rot}(z, \alpha_c)$$
$$^oT_b = \text{translate}(x_o, y_o, 0)\text{rot}(z, \alpha_o)$$

These can be combined giving the combined transformation from camera coordinates to object coordinates.

$$^oT_c = {}^oT_b{}^cT_b{}^{-1} \tag{10}$$

Let $^of_i = \lceil x_i, y_i, 0, 1\rfloor^T$ denote a homogeneous feature position in object coordinates, and let $^cz_i = \lceil x_i, y_i, 0, 1\rfloor^T$ represent an observation of f_i in camera coordinates. The vector p we are estimating is $p = [x_o, y_o, \alpha_o]^T$ of an object. We will do this by building a constraint based on correlating sensor observations z_i with features f_i.

A feature f_i appearing in the camera maps into the object frame as

$$F_i(z_i, p) = {}^cT_o{}^cz_i =$$

6

$$\begin{bmatrix} -\sin(\alpha_o)\ (y_o - y_c) - \cos(\alpha_o)\ (x_o - x_c) + \sin(\alpha_o - \alpha_c)\ y_i + \cos(\alpha_o - \alpha_c)\ x_i \\ -\cos(\alpha_o)\ (y_o - y_c) + \sin(\alpha_o)\ (x_o - x_c) + \cos(\alpha_o - \alpha_c)\ y_i - \sin(\alpha_o - \alpha_c)\ x_i \end{bmatrix} \quad (11)$$

For this single observation, we know that $F_i(z_i, p) - f_i = 0$. We can build the complete constraint for the object by letting $F(z, p) = [F_1(z_1, p), F_2(z_2, p), \ldots, F_n(z_n, p)]^T$ and $O = [f_1^T, f_2^T, \ldots, f_n^T]$. Then we we know that

$$g(z, p) = F(z, p) - O = 0 \tag{12}$$

Observation that each F_i is linear in z_i. If the z_i are independent, then the information on p resulting from each observation of z_i is independent of other observations. That is, V of Equation 7 is of the form:

$$V = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & V_n \end{pmatrix}$$

Therefore, V^{-1} will also be diagonal. The variance covariance computed in Equation 9 from independent observations will be a sum of the quadratic forms of $M_i^T V_i^{-1} M_i$. Thus, it suffices to derive the information matrix for arbitrary z_i and sum over i for the complete information matrix.

Taking partial derivative of 12 with respect to z and p gives

$$\frac{\partial g}{\partial z} = \begin{bmatrix} \cos(\alpha_o - \alpha_c) & \sin(\alpha_o - \alpha_c) \\ -\sin(\alpha_o - \alpha_c) & \cos(\alpha_o - \alpha_c) \end{bmatrix} \tag{13}$$

$$\frac{\partial g}{\partial a} =$$
$$\begin{bmatrix} -\cos(\alpha_o) & -\sin(\alpha_o) & \cos(\alpha_o)\ (y_c - y_o) + \sin(\alpha_o)\ (x_o - x_c) + \cos(\alpha_o - \alpha_c)\ y_i - \sin(\alpha_o - \alpha_c)\ x_i \\ \sin(\alpha_o) & -\cos(\alpha_o) & \sin(\alpha_o)\ (y_o - y_c) + \cos(\alpha_o)\ (x_o - x_c) - \sin(\alpha_o - \alpha_c)\ y_i - \cos(\alpha_o - \alpha_c)\ x_i \end{bmatrix} \tag{14}$$

Now, by employing Equations 8 we get the expression:

$$\Lambda_V^{-1} = \begin{bmatrix} \dfrac{\cos^2(\alpha_o - \alpha_c)}{\sigma_x^2} + \dfrac{\sin^2(\alpha_o - \alpha_c)}{\sigma_y^2} & \dfrac{\sin(2\ \alpha_o - 2\ \alpha_c)}{2\ \sigma_y^2} - \dfrac{\sin(2\ \alpha_o - 2\ \alpha_c)}{2\ \sigma_x^2} \\ \dfrac{\sin(2\ \alpha_o - 2\ \alpha_c)}{2\ \sigma_y^2} - \dfrac{\sin(2\ \alpha_o - 2\ \alpha_c)}{2\ \sigma_x^2} & \dfrac{\cos^2(\alpha_o - \alpha_c)}{\sigma_y^2} + \dfrac{\sin^2(\alpha_o - \alpha_c)}{\sigma_x^2} \end{bmatrix} \tag{15}$$

To this point, we have not considered visibility issues. However, it is clear that not all points will be visible at all times, and we can only make observations on visible points. If a point is not visible, we need to set the information associated with observing it to zero. Visibility is easily computed by the following technique:
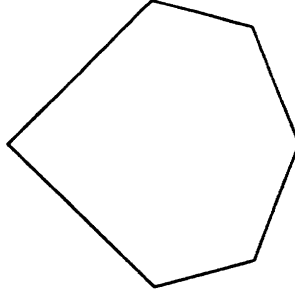
Figure 2: Sample Object

1. Project all feature points of an object to the viewing plane (line in two dimensions).

2. Compute the convex hull of the projected points.

3. Use the points forming the convex hull to divide the feature points into two classes; visible and not visible. We receive observations z_i only from those f_i that are visible.

Let us denote the set of observations resulting from visible points by

$$Z = \{z_i | z_i \text{ corresponds to a visible point}\}$$

This results in the following simple expression for the information from a single camera view:

$$\Lambda_z^{-1} = \sum_{z_i \in Z} \frac{\partial F_i(z_i, \hat{p})}{\partial z_i} \Lambda_{V_i}^{-1} \frac{\partial F_i(z_i, \hat{p})^T}{\partial z_i} \tag{16}$$

3.2 Results to Date

We have implemented these results in our laboratory. This section shows the information maps resulting from a simulated camera observing a polygon as shown in Figure 2. In order to generate these plots, we have coupled the parameters of ^bT_c by $x_c = r*\cos(\alpha_c)$ and $y_c = \cos(\alpha_c)$ and plotted with respect to r and α_c. The axis are oriented so that r is increasing to the right, and α_c is increasing to the left. The variance of observations was modeled by a constant variance on σ_y^2, and a quadratic function of the form $(x-5)^2$ on σ_x^2. This turns out to be close to models observed in depth from stereo.[1]

Figure 3 shows the determinant values resulting when σ_y^2 is equal to the minimum values achieved by σ_x^2. The maximum information values are achieved when the right hand portion of the polygon is in viewed from approximately the optimal distance. Figure 4 shows the separate components of the information matrix. The rapid variation in information values comes from the fact that Equation 16 is discontinuous due to the effects of occlusion. At every value of α_c where a point appears or disappears from view, there is an information discontinuity.
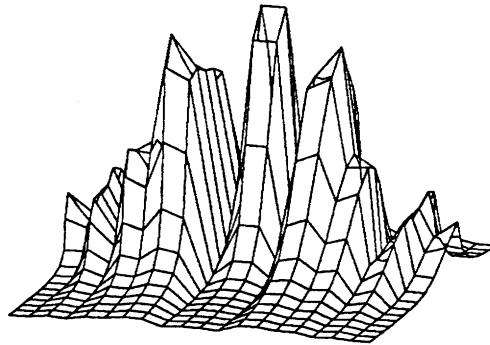
---

[1]Eric Krotkov, personal communication

8

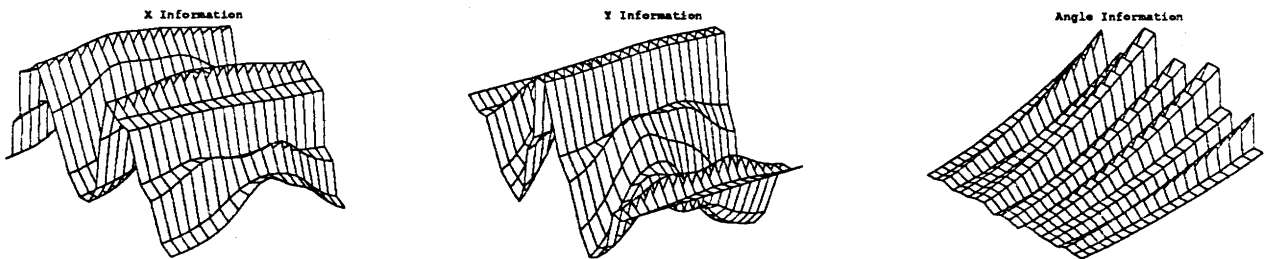Figure 3: Determinant of Λ_z^{-1} with respect to r and α_c

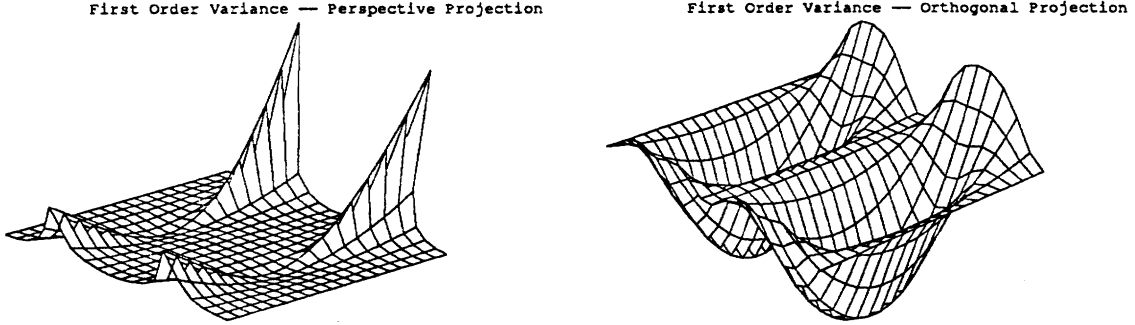Figure 4: Components of the Information Matrix

Figure 5: Plots of σ_α^2 for Orthogonal and Perspective Projections

3.3 Comparison to nonlinear Solution

For specific cases, we can evaluate the exact solution to the estimation problem and compare the results with our approximations. In the example above, the nonlinearity was in the angle α_o. If we have two points we can derive the angle of the transform by computing, via inverse tangent, the angle of the feature points; and offset that angle by the inclination of the feature relative to the object and the camera relative to the feature to get α_o. We are currently comparing the first-order approximation of $\sigma_{\alpha_o}^2$ with the actual value of $\sigma_{\alpha_o}^2$ based on this relationship. To first order, we have

$$\sigma_{\alpha_o}^2 = \frac{\cos^2(\hat{\alpha}_o + \alpha_f - \alpha_c)\sigma_y^2 + \sin^2(\hat{\alpha}_o + \alpha_f - \alpha_c)\sigma_x^2}{\|f_2 - f_1\|}, \qquad \alpha_f = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

The quantity $\hat{\alpha}_o$ represents a prior estimate of α_o. Figure 5 shows plots of variance with respect to r and α_c for orthogonal and perspective projections. In these plots, $\sigma_y^2 = 11$ and σ_x^2 follows the distance squared model. The plots we made for two feature points symmetric about a line through the origin of the object frame.

We are currently evaluating the actual values of $\sigma_{\alpha_o}^2$ under different assumed initial probability distributions. If we assume that $c_y = y_2 - y_1$ and $c_x = x_2 - x_1$ are uniformly distributed in an interval of length $2\,b_y$ and $2\,b_x$ respectively, then we can derive a closed form solution for $f_z(\cdot)$ where $z = c_y/c_x$.

$$f_z(z) = \begin{cases} \left|\min\left(\frac{c_x - b_x}{z}, c_y - b_y\right)^2 - \max\left(\frac{c_x + b_x}{z}, c_y + b_y\right)^2\right| & \text{if } z > 0 \\ \left|\min\left(\frac{c_x + b_x}{z}, c_y - b_y\right)^2 - \max\left(\frac{c_x - b_x}{z}, c_y + b_y\right)^2\right| & \text{if } z < 0 \\ \left|\min(c_y - b_y)^2 - \max(c_y + b_y)^2\right| & \text{otherwise} \end{cases}$$

Then, $\sigma_{\alpha_o}^2$ is given by the usual $E[\tan^{-1}(z)]$ which we evaluate numerically. This integral

10

is difficult evaluate since it is discontinuous an exhibits radical behavior over the integration range. Our initial results suggest that it is close to the first order approximation.

If we assume that $c_y = y_2 - y_1$ and $c_x = x_2 - x_1$ are normally distributed with unequal variance and non-zero mean, then we cannot even derive a closed form solution for $f_z(\cdot)$ where $z = c_y / c_x$ in general. It is known that, for zero mean processes, the final distribution is

$$f_\alpha(\cdot) = \frac{b}{\pi} \frac{b}{b^2 \cos^2(\alpha) + sin^2(\alpha)} \qquad \text{where } b = \sqrt{\frac{\sigma_y^2}{\sigma_x^2}}$$

Even in this case, σ_α^2 is not analytically derivable. For the case of non-zero mean, we must do a numerical integration to derive the distribution

$$f_z(z) = \int_{-\infty}^{\infty} |y| \, e^{-1/2 \left(\frac{(zy - c_x)^2}{\sigma_z^2} + \frac{(y - c_y)^2}{\sigma_y^2} \right)} dy$$

Then, to evaluate $E[\tan^{-1}(z)]$, we must again numerically evaluate

$$\int_{-\infty}^{\infty} \tan^{-1}(z) \, f_z(z) dz$$

We are currently checking our implementation of this integration. Our initial results suggest that the first order approximation is not a good approximation of the correct value.

4   Robustness Issues

It should be noted that the resulting information maps are really random variables. That is, the variance is conditioned on observations. Moreover, the computed information maps depend on a linearization. The linearization is done *at a given point*. This technique depends on having a relatively dependable prior to furnish the point to linearize about. If the previous estimate is bad, then we can expect that the linearization, and hence the information maps, diverge substantially from the true values. There are a number of approaches to solving this problem. For instance, we could go to a second or higher order approximation, but we loose the advantage of a simple linear system.

One approach we are considering pursuing is to apply game-theoretic techniques to the problem. Recall from Equation 1 that our measurement system took the form

$$z = H(u, p) + V(u)$$

In order to linearize this, we were forced to pick a prior point estimate for p. Rather than do this, we can instead think of a relatively bad p so that whatever action we take is safe. In this case, we can state our estimation problem in the following form:

$$\min_{A} \max_{H} E\|Az - a\|^2 \qquad H \in \mathcal{H} \tag{17}$$

We specify a set of possible H's, $\mathcal{H}$ and choose our estimation procedure based on the worst member of that set. Some general results in game theory assure us that there is a *single* best A, but possibly multiple (though a finite number) of H's with an associated distribution. In the case of multiple values for H, we will use the distribution as though it were our prior on H and solve the estimation problem. Application of these techniques should yield robust information maps.

5   Conclusions and Discussion

We have presented results pertaining to the prediction of information from sensor systems. We consider the sensor as a controllable measurement system observing the environment. These observation are related to the parameters of interest via general constraints. We apply linearization techniques to derive a linear observation system, and look at the variance-covariance matrices resulting from mean-square estimation procedures. These results are extended to account for occlusion and missing observations.

Concurrent with this, we are attempting to evaluate the robustness of linearized constraints. In general, linearization relies on having a reasonable prior estimate to linearize about, and smoothness of the function being linearized. We are investigating robust techniques for estimation base on game theory. Our initial investigations seem promising.

We plan to use information maps for control of sensors. Several interesting questions arise due to the scope limitations of sensors. For instance, it is possible the information of interest is not observable from the current vantage points. What policy should we employ to make it observable? If the information is observable, then, by the law of large numbers, we can always get good information by taking a sufficient number of observations. What if we have time and processor constraints? Should we move? Where should we move? We foresee addressing all of these questions.

Acknowledgements: The author would like to thank Max Mintz for his help and guidance in doing this research.

References

[1] R. Bajcsy. Active perception vs. passive perception. In *Third IEEE Workshop on Vision*, pages 55–59, Bellaire, 1985.

[2] G. Hager. *Active Reduction of Uncertainty in Multi-Sensor Systems.* Grasp Lab 79, University of Pennsylvania, Philadelphia, PA, September 1986. Author's Dissertation Proposal, also MS-CIS-86-76.

[3] I. Lewis Meier, J. Peschon, and R. M. Dressler. Optimal control of meaurement subsystems. *IEEE Transactions on Automatic Control,* AC-12(5):528–536, October 1967.

[4] P. Müller and H. Weber. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica,* 8:237–246, 1972.

[5] R. Paul. *Robot Manipulators.* MIT Press, Cambridge, 1981.