In *Proc. of the 1994 Conf. on Computer Vision and Pattern Recognition (CVPR94)*, June 20-24, 1994, Seattle, Washington.

The Resolvability Ellipsoid for Visual Servoing

Bradley J. Nelson Pradeep K. Khosla

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3891

Abstract

This paper introduces a sensor placement measure called resolvability. The measure provides a technique for estimating the relative ability of various visual sensors, including monocular systems, stereo pairs, multi-baseline stereo systems, and 3D rangefinders, to accurately control visually manipulated objects. The resolvability ellipsoid illustrates the directional nature of resolvability, and can be used to direct camera motion and adjust camera intrinsic parameters in real-time so that the servoing accuracy of the visual servoing system improves with camera-lens motion. The Jacobian mapping from task space to sensor space is derived for a monocular system, a stereo pair with parallel optical axes, and a stereo pair with perpendicular optical axes. Resolvability ellipsoids based on these mappings for various sensor configurations are presented. Visual servoing experiments demonstrate that resolvability can be used to direct camera-lens motion in order to increase the ability of a visually servoed manipulator to precisely servo objects.

1. Introduction

In order to effectively use visual feedback to perform robotic tasks, many researchers have recognized that the placement of the sensor relative to the task is an important consideration, for example [1], [8], and [9]. To determine optimal sensor configurations, various sensor placement criteria are required. This paper presents a new sensor placement measure called *resolvability*. The measure provides a technique for estimating the relative ability of various visual sensor systems, including single camera systems, stereo pairs, multi-baseline stereo systems, and 3D rangefinders, to accurately control visually manipulated objects and to provide spatially accurate data on objects of interest.

Sensor resolution has been considered in the past as a criterion for sensor planning [1], [7], and [9]. These efforts concern static camera systems in which a required spatial resolution is known and a single camera placement is desired. In [2], a study of stereo, vergence, and focus cues for determining range is described in which the performance of each cue for determining range accuracy is characterized. This characterization can be used to control camera parameters in order to improve the accuracy of range estimates. Our resolvability approach can be used for determining the ability of a visually servoed manipulator to accurately resolve the position and orientation of objects along all six degrees of freedom. An ellipsoidal representation of resolvability illustrates its directional nature. Intrinsic and extrinsic camera-lens parameters can be actively controlled using resolvability in conjunction with other sensor placement criteria to improve the accuracy of visual control. Resolvability can also be used for static sensor placement for object modeling or visual servoing.

2. Task space → sensor space

Resolvability is a function of the Jacobian mapping from task space to sensor space. For any visual sensor system, we desire an equation of the form

$$\dot{\boldsymbol{x}}_S = \boldsymbol{J}(\phi)\dot{\boldsymbol{x}}_T \tag{1}$$

where $\dot{\boldsymbol{x}}_S$ is a velocity vector in sensor space and $\dot{\boldsymbol{x}}_T$ is a velocity vector in task space. $\boldsymbol{J}(\phi)$ is the Jacobian matrix and is a function of the extrinsic and intrinsic parameters of the visual sensor as well as the number of features tracked and their locations on the image plane.

By performing a singular value decomposition [3] on the task space to sensor space Jacobian, and analyzing the singular values and the eigenvectors of $\boldsymbol{J}^T\boldsymbol{J}$ which result from the decomposition, the directional properties of the ability of the sensor to resolve positions and orientations becomes apparent. These directional properties can be represented by the resolvability ellipsoid, if we assume that the object of interest has an equal ability to translate and rotate about all of its cartesian axes. The following sections briefly describe the derivation of the Jacobian mapping.

2.1. Jacobian mapping for monocular systems

A pinhole camera model shown in Figure 1 is used to model the camera-lens system. A feature on an object at ${}^C\boldsymbol{P}$ with coordinates (X_C, Y_C, Z_C) in the camera frame projects onto the camera's image plane at

$$x_S = \frac{fX_C}{s_x Z_C} \qquad y_S = \frac{fY_C}{s_y Z_C} \tag{2}$$

where f is the focal length of the lens and s_x and s_y are the horizontal and vertical dimensions of the pixels on the CCD array.

For visually servoing a manipulator holding an object, the velocity of ${}^C\boldsymbol{P}$ is induced relative to the tool frame {T} according to

$${}^C\dot{\boldsymbol{P}} = {}^C_T\boldsymbol{R}\left({}^T\boldsymbol{V} + {}^T\dot{\boldsymbol{P}} + {}^T\Omega \times {}^T\boldsymbol{P}\right) \tag{3}$$

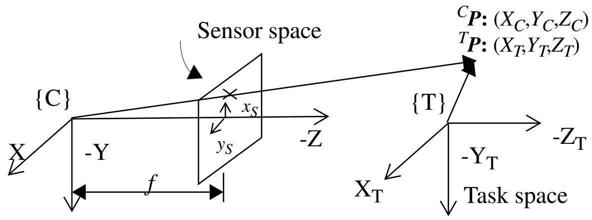

Figure 1. Task frame-camera frame definitions.

where ${}^{T}\boldsymbol{V} = \begin{bmatrix}\dot{x}_T & \dot{y}_T & \dot{z}_T\end{bmatrix}^T$ and ${}^{T}\Omega = \begin{bmatrix}\omega_{x_T} & \omega_{y_T} & \omega_{z_T}\end{bmatrix}^T$ are the translational and rotational velocities of the task frame with respect to itself. These are manipulator end-effector velocities that can be commanded. By combining (3) with image plane optical flow described by the time derivatives of (2), the entire Jacobian transformation for a single feature from task space to sensor space can be written in the form

$$\begin{bmatrix}\dot{x}_S \\ \dot{y}_S\end{bmatrix} = \begin{bmatrix} \frac{f}{s_x Z_C} & 0 & -\frac{x_S}{Z_C} & -\frac{Y_T x_S}{Z_C} & \left[\frac{fZ_T}{s_x Z_C} + \frac{X_T x_S}{Z_C}\right] & -\frac{fY_T}{s_x Z_C} \\ 0 & \frac{f}{s_y Z_C} & -\frac{y_S}{Z_C} & -\left[\frac{fZ_T}{s_y Z_C} + \frac{Y_T y_S}{Z_C}\right] & \frac{X_T y_S}{Z_C} & \frac{fX_T}{s_y Z_C} \end{bmatrix} \begin{bmatrix}\dot{x}_T \\ \dot{y}_T \\ \dot{z}_T \\ \omega_{x_T} \\ \omega_{y_T} \\ \omega_{z_T}\end{bmatrix} \tag{4}$$

The parameters of the Jacobian are given by $\phi = (f, s_x, s_y, x_S, y_S, Z_C, X_T, Y_T, Z_T)$. Generally, several features on an object are tracked. For n feature points, the Jacobian is of the form

$$\boldsymbol{J} = \begin{bmatrix}\boldsymbol{J}_1 & \dots & \boldsymbol{J}_n\end{bmatrix}^T \tag{5}$$

where $\boldsymbol{J}_i$ is the Jacobian matrix for each feature given by the 2x6 matrix in (4).

2.2. Jacobian mapping for binocular systems with parallel optical axes

For a binocular system with parallel optical axes, the locations of the camera frame and task frame are shown in Figure 2. The camera model is represented by

$$x_{Sl} = \frac{f\left(X_C + \frac{b}{2}\right)}{s_x Z_C} \qquad y_{Sl} = \frac{fY_C}{s_y Z_C} \tag{6}$$

$$x_{Sr} = \frac{f\left(X_C - \frac{b}{2}\right)}{s_x Z_C} \qquad y_{Sr} = \frac{fY_C}{s_y Z_C} \tag{7}$$

where b is the length of the baseline of the cameras, and it is assumed that f, s_x, and s_y are the same for both cameras.Through a similar derivation as in Section 2.1, the mapping from task space velocity to sensor space velocity can be written as

$$\begin{bmatrix} \frac{d}{b} & 0 & -x_{Sl}\frac{s_x d}{fb} & -x_{Sl}\frac{s_x d}{fb}Y_T & \left[\frac{d}{b}Z_T + x_{Sl}\frac{s_x d}{fb}\left(X_T + \frac{b}{2}\right)\right] & -\frac{d}{b}Y_T \\ 0 & \frac{s_x d}{s_y b} & -y_{Sl}\frac{s_x d}{fb} & -\left[\frac{d}{s_y b}Z_T + y_{Sl}\frac{s_x d}{fb}Y_T\right] & y_{Sl}\frac{s_x d}{fb}\left(X_T + \frac{b}{2}\right) & \frac{s_x d}{s_y b}\left(X_T + \frac{b}{2}\right) \\ \frac{d}{b} & 0 & -x_{Sr}\frac{s_x d}{fb} & -x_{Sr}\frac{s_x d}{fb}Y_T & \left[\frac{d}{b}Z_T + x_{Sr}\frac{s_x d}{fb}\left(X_T - \frac{b}{2}\right)\right] & -\frac{d}{b}Y_T \\ 0 & \frac{s_x d}{s_y b} & -y_{Sr}\frac{s_x d}{fb} & -\left[\frac{s_x d}{s_y b}Z_T + y_{Sr}\frac{s_x d}{fb}Y_T\right] & y_{Sr}\frac{s_x d}{fb}\left(X_T - \frac{b}{2}\right) & \frac{s_x d}{s_y b}\left(X_T - \frac{b}{2}\right) \end{bmatrix} \tag{8}$$

where $d = x_{Sl} - x_{Sr}$ is the disparity of each corresponding feature point. The sensor space vector contains four terms representing the optical flow of the feature in both the left and right images. The task space vector is the same as in (4).

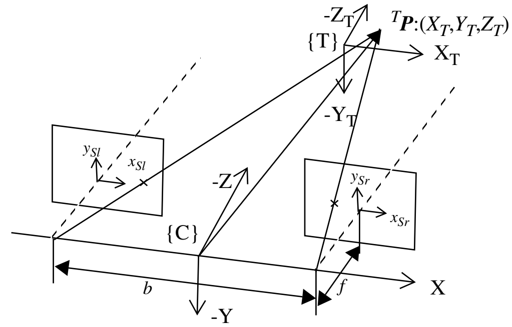

Figure 2. Task frame-camera frame definitions.

2.3. Jacobian mapping for binocular systems with orthogonal optical axes

An orthogonal stereo pair is shown in Figure 3. If the axes are aligned as shown in the figure, the Jacobian mapping from task space to sensor space can be written as

$$\begin{bmatrix} \frac{f}{s_x Z_{Cl}} & 0 & -\frac{x_{Sl}}{Z_{Cl}} & -\frac{x_{Sl}Y_T}{Z_{Cl}} & \left[\frac{fZ_T}{s_x Z_{Cl}} + \frac{x_{Sl}X_T}{Z_{Cl}}\right] & -\frac{fY_T}{s_x Z_{Cl}} \\ 0 & \frac{f}{s_y Z_{Cl}} & -\frac{y_{sl}}{Z_{Cl}} & -\left[\frac{fZ_T}{s_y Z_{Cl}} + \frac{y_{Sl}Y_T}{Z_{Cl}}\right] & \frac{y_{Sl}X_T}{Z_{Cl}} & \frac{fX_T}{s_y Z_{Cl}} \\ \frac{x_{Sr}}{Z_{Cr}} & 0 & \frac{f}{s_x Z_{Cr}} & \frac{fY_T}{s_x Z_{Cr}} & \frac{x_{Sr}Z_T}{Z_{Cr}} - \frac{fX_T}{s_x Z_{Cr}} & \frac{x_{Sr}Y_T}{Z_{Cr}} \\ \frac{y_{Sr}}{Z_{Cr}} & \frac{f}{s_y Z_{Cr}} & 0 & -\frac{fZ_T}{s_y Z_{Cr}} & \frac{y_{Sr}Z_T}{Z_{Cr}} & \frac{y_{Sr}Y_T}{Z_{Cr}} - \frac{fX_T}{s_y Z_{Cr}} \end{bmatrix} \tag{9}$$

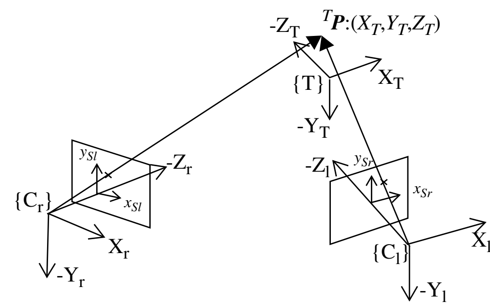

Figure 3. Task frame-camera frame definitions.

3. Resolvability Ellipsoids

In this section, resolvability ellipsoids for several different camera-lens configurations are presented. To show the ellipsoidal representation, the Jacobian mapping is decomposed into two mappings, one representing translational components and one representing rotational components.

In Figures 4 and 5, ellipsoids for a monocular system are shown in which the two examples have the same magnification (f/Z_C), but the object is located at different depths. Figure 6 is a plot of resolvability in depth versus depth and focal length. From the plot one can observe that progressively smaller depths have progressively larger effects on resolvability in depth, while focal length tends to affect

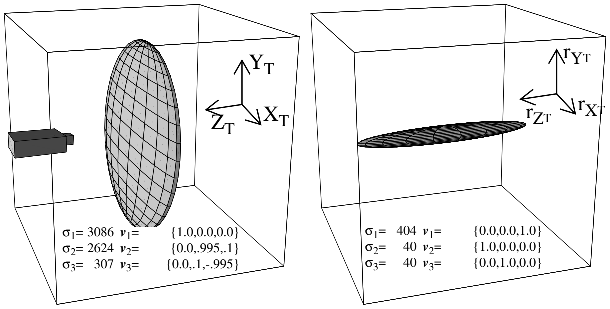

Figure 4. Resolvability Ellipsoids: monocular system, f=24mm, depth=1.0m, 2 features located in the task frame at (0.1m,0.1m,0) and (-0.1m,0.1m,0).

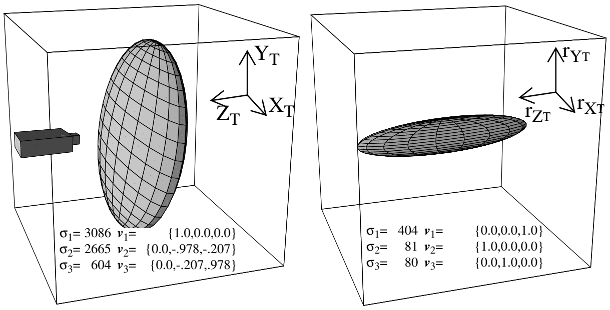

Figure 5. Resolvability Ellipsoids: monocular system, f=12mm, depth=0.5m, 2 features located in the task frame at (0.1m,0.1m,0) and (-0.1m,0.1m,0).

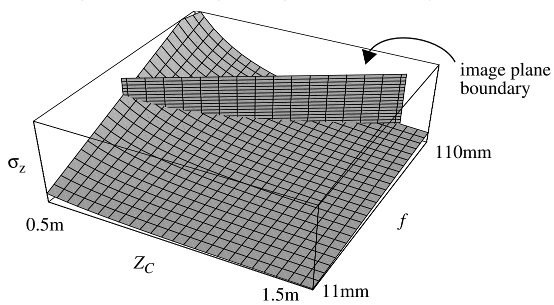

Figure 6. Resolvability of depth versus depth of object and focal length for two features located in the task frame at (0.05m,0,0) and (-0.05m,0,0).

depth resolvability more linearly. In practice, depth becomes limited by the depth-of-field of the lens, and a trade-off must be made between focal length, depth, depth-of-field, and field-of-view. Figure 7 shows the resolvability about the optical axes versus the position at which an object is observed on the image plane. The closer the object's projection falls to the boundary of the image plane, the greater the resolvability about the optical axis.

Figure 8 shows resolvability ellipsoids for a binocular system tracking a single feature. Depth can be resolved using a single feature, but not accurately relative to directions parallel to the image plane. Figure 9 shows a plot of resolvability in depth versus baseline and depth. This plot demonstrates that reducing depth is preferable to extending the baseline to improve resolvability in depth.

The resolvability ellipsoids for a binocular system with orthogonal optical axes are shown in Figure 10. The con-

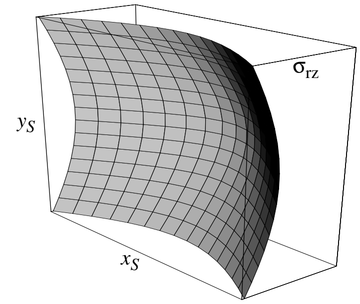

Figure 7. Resolvability in orientation about Z versus center of object projection onto the image plane.

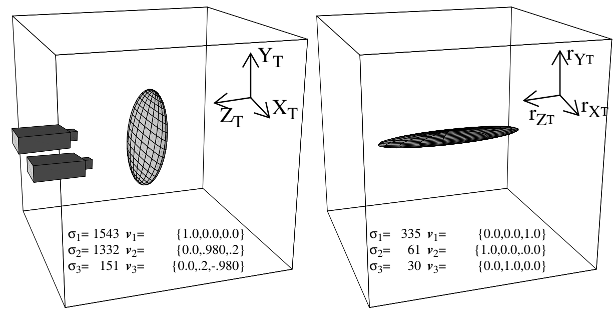

Figure 8. Resolvability Ellipsoids: stereo pair-parallel optical axes, f=12mm, b=20cm, depth=1.0m, 1 feature located in the task frame at (0,0.2m,0).

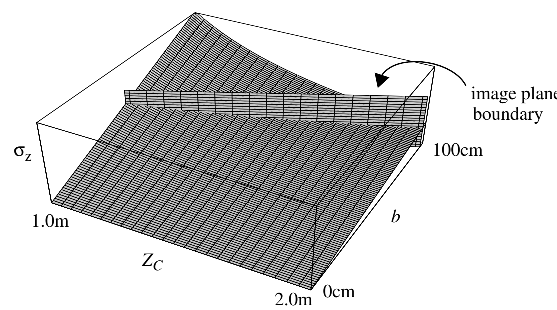

Figure 9. Resolvability in depth versus baseline length and depth of object for a stereo pair, parallel optical axes, f=12mm, and a single feature located at the origin of the task frame.

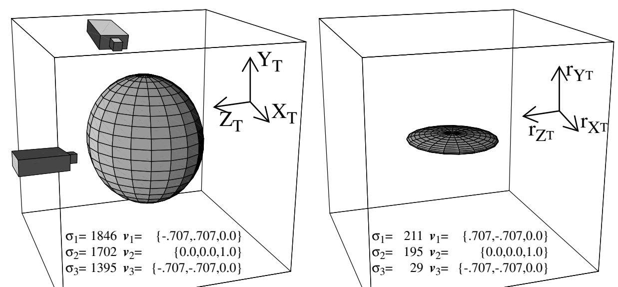

Figure 10. Resolvability Ellipsoids: stereo pair-perpendicular optical axes, f=12mm, depth=1.0m, 2 features located in the task frame at (-0.1m,0.1m,0), and (0.1m,-0.1m,-0.1m).

figuration provides a very well conditioned Jacobian mapping from task space to sensor space, although resolvability about Y_T is still relatively low.

4. Directing Camera-Lens Motion Using the Resolvability Ellipsoid

Our primary interest in resolvability is for actively guiding camera-lens motion while performing visually servoed manipulation tasks. A gradient in camera-lens parameter space can be determined which directs camera-lens motion towards configurations with improved resolvability. For example, in order to determine motions which will increase resolvability along the optical axis, the gradient of σ_Z with respect to the camera-lens parameter space is calculated, where σ_Z represents the singular value which corresponds to the eigenvector along the optical axis. The gradient is written as

$$\nabla_{\phi}\sigma_Z = \left[\frac{\partial\sigma_Z}{\partial x_C}\ \frac{\partial\sigma_Z}{\partial y_C}\ \frac{\partial\sigma_Z}{\partial z_C}\ \frac{\partial\sigma_Z}{\partial r_{X_C}}\ \frac{\partial\sigma_Z}{\partial r_{Y_C}}\ \frac{\partial\sigma_Z}{\partial r_{Z_C}}\ \frac{\partial\sigma_Z}{\partial f}\right] \qquad (10)$$

The individual components of $\nabla_{\phi}\sigma_Z$ are calculated numerically. Camera-lens motion is then directed along $\nabla_{\phi}\sigma_Z$ in order to increase σ_Z subject to other sensor placement criteria such as depth-of-field and field-of-view measures, which simultaneously affect camera-lens motion. A technique for integrating this gradient into a visual tracking control law for a hand/eye system can be found in [4].

5. Experimental Results

Visual servoing experiments were performed that demonstrate resolvability. Controlled active vision [6] was used to derive a visual servoing control law of the form

$$\boldsymbol{u}(k) = -\left(\boldsymbol{J}^T(k)\boldsymbol{Q}\boldsymbol{J}(k) + \boldsymbol{R}\right)^{-1}\boldsymbol{J}^T(k)\boldsymbol{Q}\,[\boldsymbol{x}(k) - \boldsymbol{x}_D(k+1)] \qquad (11)$$

where $\boldsymbol{x}(k) \in R^{2M}$, T is the sampling period of the vision system, $\boldsymbol{u}(k) = \left[\dot{x}_T\ \dot{y}_T\ \dot{z}_T\ \omega_{x_T}\ \omega_{y_T}\ \omega_{z_T}\right]^T$ is the commanded manipulator end-effector velocity, and M is the number of features being tracked. $\boldsymbol{Q}$ and $\boldsymbol{R}$ are weighting matrices and allow the user to place a varying emphasis on the feature error and the control input.

Figure 11 shows results of an experiment in which an object was placed in the gripper of a manipulator and visually servoed so that the length of the projection of the object decreased by 10 pixels. To achieve this, motion along Z is required. The greater the resolvability, the smaller the required motion along Z. The plot shows that as resolvability along Z increases with decreasing initial depth, the smaller the required motion along Z. A more complete explanation of the experiments performed can be found in [5].

6. Conclusion

The directional nature of the resolvability ellipsoid makes this sensor placement criterion particularly useful for guiding visual sensor motion in real-time, or as an aid in determining the placement of static sensors. In this paper we have shown that resolvability can effectively represent the ability of different sensor configurations to resolve the

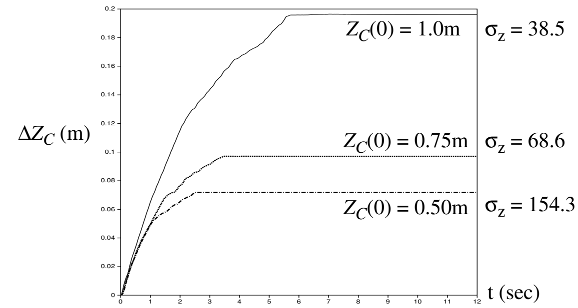

Figure 11. Change in object depth required to affect a 10 pixel change in projected object length for three different initial depths and a focal length of 15mm.

position and orientation of objects being observed. For visual servoing, this concept directly relates to the accuracy with which a manipulator can move an object to some desired goal position and orientation. Experimental results demonstrate that resolvability can be used to direct camera-lens motion in particular directions in order to increase the ability of a visually servoed manipulator to perform precise manipulation tasks.

Acknowledgments

This research was supported by the U.S. Army Research Office through Grant Number DAAL03-91-G-0272 and by Sandia National Laboratories through Contract Number AC-3752D.

References

[1] C.K. Cowan and P.D. Kovesi, "Automatic Sensor Placement from Vision Task Requirements," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 407-416, 1988.

[2] S. Das and N. Ahuja, "A Comparative Study of Stereo, Vergence, and Focus as Depth Cues for Active Vision," in *Proc. of the 1992 Computer Vision and Pattern Recognition Conference (CVPR-92)*, 194-199, 1993.

[3] V.C. Klema and A.J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," in *IEEE Transactions on Automatic Control*, 25(2), 164-176, 1980.

[4] B. Nelson and P.K. Khosla, "Integrating Sensor Placement and Visual Tracking Strategies," in *EXPERIMENTAL ROBOTICS III: The Third International Symposium, Kyoto, October 28-30 1993*, eds. T. Yoshikawa and F. Miyazaki, Springer-Verlag London Ltd., London, 1993.

[5] B. Nelson and P.K. Khosla, "The Resolvability Ellipsoid for Visually Guided Manipulation," Technical Report CMU-RI-TR-93-28, The Robotics Institute, Carnegie Mellon University, 1993.

[6] N.P. Papanikolopoulos, B. Nelson, and P.K. Khosla, "Monocular 3-D Visual Tracking of a Moving Target by an Eye-in-Hand Robotic System," in *Proc. of the 31st IEEE Conf. on Decision and Control (31stCDC)*, 3805-3810, 1992.

[7] K. Tarabanis, R.Y. Tsai, and P.K. Allen, "Satisfying the Resolution Constraint in the "MVP" Machine Vision Planning System," in *Proc. of the 1990 Darpa Image Understanding Workshop*, 850-860, 1990.

[8] K. Tarabanis, R.Y. Tsai, and P.K. Allen, "Automated Sensor Planning for Robotic Vision Tasks," in *Proc. of the 1991 IEEE Int. Conf. on Robotics and Automation*, 76-82, 1991.

[9] S. Yi, R.M. Haralick, and L.G. Shapiro, "Automatic Sensor and Light Positioning for Machine Vision," in *Proc. of the 10th Int. Conf. on Pattern Recognition*, 55-59, 1990.