Proceedings of the 1991 IEEE
International Conference on Robotics and Automation
Sacramento, California - April 1991

Automated sensor planning for robotic vision tasks

Konstantinos Tarabanis*
Computer Science Department
Columbia University
New York, New York 10027

Roger Y. Tsai
Manufacturing Research
IBM T.J. Watson Research Center
Yorktown Heights, New York 10598

Peter K. Allen
Computer Science Department
Columbia University
New York, New York 10027

Abstract

In this paper, we present a method to determine viewpoints for a robotic vision system for which object features of interest will simultaneously be visible, inside the field-of-view, in-focus and magnified as required. As part of our previous work, we had analytically characterized the domain of admissible camera locations, orientations and optical settings for which each of the above feature detectability requirements is satisfied separately. In this paper, we present a technique that poses the problem in an optimization setting in order to determine viewpoints that satisfy all requirements simultaneously and with a margin. The formulation and results of the optimization are shown, as well as, experimental results in which a robot vision system is positioned and its lens is set according to this method. Camera views are taken from the computed viewpoints in order to verify that all feature detectability requirements are indeed satisfied.

1 PROBLEM

In a general sense, sensor planning can be defined to embody a number of areas of robotics that have been studied extensively in the past. For instance, the general problem of *task planning* and its component areas of *motion planning*, *grasp planning* and *assembly planning* can be seen as different facets of the sensor planning problem. Even when limiting the problem to planning of vision sensors alone, two distinct areas can be observed based on the vision task that is to be achieved. One area is concerned with developing sensing strategies for the tasks of *object recognition*, *reconstruction* or *localization.* That is, it is concerned with choosing sensing operations and sensing positions, or determining object features, that will prove most useful when trying to identify or reconstruct an object or determine its pose. Along orthogonal lines, there has been work in which the sensor planning problem is posed in a *decision theoretic framework*. That is, statistical decision theory is used to determine optimal sensing locations for performing a task, with sensors being modeled as noisy information sources [1, 4]. The other area considers the vision task of *object feature detection.* That is, it determines sensing parameters for which particular features on a known object satisfy particular constraints in the image (e.g. the feature is visible, in-focus and magnified as required) [2, 3], [7, 8], [18], [10, 11, 13, 14, 16, 17]. In this paper, we present work in this latter area.

2 OVERVIEW OF OUR APPROACH

We are developing a *model-based and task-driven vision system* MVP, (Machine Vision Planner), that automatically *plans* vision sensor parameters so that task requirements, common to most industrial machine vision applications, are satisfied. Methods are being developed that take as input the object geometry information, as well as models of the camera and lens, and determine camera poses and optical settings for which features of interest of polyhedral objects are:

- visible (occlusion-free positions of the sensor),
- contained entirely in the sensor field-of-view,
- in-focus,
- resolvable by the sensor to a given specification.

These task requirements all determine *feature detectability* and therefore are fairly generic for most vision tasks.

The planning techniques that are being developed are used in a robotic vision system, in which a camera and light source are mounted on two robot manipulators. These robot manipulators can position and orient the camera and light source, while the camera optics can be

* this author was supported in part by DARPA contract N00039-84-C-0165 and in part by Manufacturing Research, IBM T.J. Watson Research Center.

controlled, either by manually exchanging fixed focal length lenses or by controlling the zoom, focus and aperture settings of a programmable zoom lens [16]. In addition, the robotic vision system has access to a geometry database of objects to be observed.

The sensor parameters that are planned are geometric and optical in nature. The geometric parameters are the three positional degrees of freedom of the sensor- $\vec{r}_O(x, y, z)$-, and the two orientational degrees of freedom, *pan and tilt angles*, described by a unit vector $\vec{v}$ along the viewing direction (rotation around the optical axis is ignored). On the other hand, there are three optical parameters, namely, the back nodal point[1] to image plane distance d, the focal length f and the aperture of the lens a. Thus, planning is done in *eight-dimensional imaging space* [9] and a point in this space is a *generalized viewpoint* $V(\vec{r}_O, \vec{v}, d, f, a)$.

Using concepts from geometry, illumination and optics, each task requirement is modeled by an equivalent *analytical relationship*, which in turn is satisfied in a domain of admissible values in the space of parameters to be planned. For each constraint, the admissible domain for sensor placement and setting is a region in eight-dimensional imaging space bounded by the hypersurfaces that are characterized by these analytical relationships.

These component admissible domains obtained for each task requirement need to be combined in order to find parameter values that satisfy all constraints simultaneously. If the entire admissible domain of parameter values is sought, then, in principle, the combination of these component solutions involves *intersecting* eight-dimensional regions in order to determine solutions that are admissible to all constraints. However such intersections are still research problems in and of themselves and therefore more practical techniques need to be developed.

For this reason, the problem has been posed in an *optimization* setting, in which a globally admissible eight-dimensional viewpoint is sought that is "*central*" to the admissible domain, that is, far from the bounding hypersurfaces described by the constraint equations. Such a generalized viewpoint is desirable, since it is robust in the event of inaccuracy (e.g. due to sensor noise) of either sensor placement or setting. The analytical relationships for each task constraint provide the constraints for the optimization, while the objective function is chosen so as to characterize the "distance" between a viewpoint and these bounding hypersurfaces.

Once a "central" generalized viewpoint is determined from the optimization, it needs to be realized in the actual sensor setup. In order to achieve these planned sensor parameter values, a mapping needs to be established between the planned parameters (e.g. camera pose and optical settings) and the parameters that can be controlled (e.g. end effector pose, zoom and focus settings). This mapping between the two parameter spaces is provided by the *calibration models*. These models embody knowledge of the geometric relationships of the manipulator, the sensor and illuminator, as well as the optical relationships of the lenses.

3 THE VIEWPOINT LOCI

In this section, we briefly describe the determination of the loci of admissible generalized viewpoints for each constraint separately. Details and derivations of the following can be found in [13].

3.1 FEATURE VISIBILITY CONSTRAINT

The domain of admissible sensor locations is first limited to regions in three-dimensional space from where the features to be observed are *visible*. We have developed a technique that generates regions in three-dimensional space from where features of interest on an object can be viewed in their entirety without being obstructed by the object itself (i.e. self-occlusion). The features to be observed can be of any polyhedral type: a point, line-segment or face (including concave faces). The details of this technique are described in [12].

The visibility planning algorithm first considers a sufficient subset of the faces of the observed polyhedron as polygons in three-dimensional space, possibly concave, that are potentially occluding the feature to be observed. The algorithm then determines the three-dimensional occluded regions between these occluding polygons and the target feature. These individual occluded regions of the faces of the polyhedron are then unioned to generate the occluded region of the polyhedron as a whole. The complement of the occluded region is the visibility region, from where the entire target can be viewed (see Figure 1).

3.2 FEATURE RESOLUTION CONSTRAINT

Pixel resolution is used to indicate the approximate size of the smallest scene feature which can be seen by the vision system. In many machine vision tasks it is required that a particular unit feature size on an object appear as a minimum number of picture elements on a sensor. This feature resolution constraint can be satisfied by properly

1 Nodal points are points on the optical axis whose properties are such that any ray passing through the front nodal point emerges from the back nodal point in the direction parallel to that of the original ray. For the case of a thin lens the nodal points coincide at the perspective center.

selecting the image sensor (e.g. pixel size), as well as by carefully planning its placement and settings. The objective of sensor planning for the feature resolution constraint is to determine the sensor parameters that achieve this resolution. The type of features that are considered for the resolution constraint are line-segments, thus including feature edges or linear features of interest (e.g. width between two edges). We have developed a method to plan the camera pose and the optical settings of a lens, so that chosen features can be resolved to meet a given specification, for instance, feature A_iB_i has an image of length that is at least equal to w. This locus of resolution satisfying generalized viewpoints is described in vector form by the following formula:

$$g_1: \quad \frac{\|\vec{v}\times[\vec{e_i}\times(\vec{r_O}-\vec{r_A})]\|}{[(\vec{r_A}-\vec{r_O})\bullet\vec{v}][(\vec{r_B}-\vec{r_O})\bullet\vec{v}]} - \frac{w}{dl} \geq 0$$

where

$\vec{r_O}$, $\vec{r_A}$, $\vec{r_B}$, are the position vectors of the front nodal point of the lens and the feature vertices with respect to the object world coordinate system O_W,

d is the effective focal length, (i.e. the distance from the back nodal point of the lens to the image plane),

$\vec{v}$ is the unit vector along the optical axis in the viewing direction,

$\vec{e_i}$ is the unit vector along the feature edge,

f is the intrinsic focal length of the lens, that is, the focal length of the lens for an object at infinity,

l_i is the length of the minimum feature to be resolved, and,

w is the required length of $A'_iB'_i$, the image of A_iB_i.

3.3 THE DEPTH-OF-FIELD CONSTRAINT

When planning camera placement and lens settings so that all features of interest on an object are in-focus simultaneously, this corresponds to determining the locus of generalized viewpoints for which the feature points that are farthest and nearest with respect to each viewing direction lie within the range described by the depth-of-field. It can be shown that the region that satisfies the depth-of-field constraint is given in vector form by the following formulas:

$$g_{2a}: \quad \|(\vec{r}_O - \vec{r}_c)\bullet\vec{v}\| - D_2 \geq 0$$

$$g_{2b}: \quad D_1 - \|(\vec{r}_O - \vec{r}_f)\bullet\vec{v}\| \geq 0$$

where

D_1 is the far limit of the depth-of-field given by [5] $D_1 = \dfrac{Daf}{af - c(D-f)}$,

D_2 is the near limit of the depth-of-field given by [5] $D_2 = \dfrac{Daf}{af + c(D-f)}$,

D is the focus distance given by $D = \dfrac{1}{(1/f - 1/d)}$,

$\vec{r}_f$ is the position vector of the farthest feature vertex from the front nodal point of the lens along the viewing direction,

$\vec{r}_c$ is the position vector of the closest feature vertex from the front nodal point of the lens along the viewing direction,

and all other variables are as defined in previous sections.

3.4 THE FIELD-OF-VIEW CONSTRAINT

In section 3.1 it was implicitly assumed that there were no field-of-view limitations, that is, the sensor had a 180 degree field-of-view angle and therefore orientation of the sensor was immaterial, provided that the features to be observed were in the half-space associated with the front of the camera. For a CCD camera, the field-of-view is generally limited by the minimum dimension $I_{\min}$ corresponding to the active sensor area in the image plane-that is, the sensor plane is the field-stop[2] of the system. Any observed feature must project onto this limited image plane, otherwise it will either be totally outside the field-of-view or truncated. It can be shown that the relationship describing the field-of-view satisfying locus of generalized viewpoints is given in vector form by the following formula:

$$g_3: \quad (\vec{r}_K - \vec{r}_O)\bullet\vec{v} - \cos(\tfrac{\alpha}{2})\|\vec{r}_K - \vec{r}_O\| \geq 0$$

where

α is the field-of-view angle and is given by $\alpha = 2\tan^{-1}(I_{\min}/2d)$,

$I_{\min}$ is the minimum dimension of the sensor plane,

2 the field-stop is the stop in the optical system that limits the field-of-view.

$\vec{r_C}$ is the position vector of the center of the sphere circumscribing the object features,

R_f is the radius of the sphere circumscribing all the object features,

$\vec{r_K}$ is the position vector given by $\vec{r}_K = \vec{r}_C - R_O\vec{v}$,

R_O $= R_f/(\sin \alpha/2)$,

and all other variables are as defined in previous sections.

4 CONSTRAINT MERGING

Constraint merging is formulated as a constrained optimization problem. The constraints of the optimization problem include the feature detectability requirements, g_i, $i = 1,2a,2b,3$, that were presented in previous sections. An additional optimization constraint, g_5, expresses the unit vector condition for the viewing vector $\vec{v}$ that appears in relationships g_1, g_{2a}, g_{2b} and g_3. That is, g_5: $\|\vec{v}\|^2 - 1 = 0$. This constraint g_5 is an equality, whereas g_1, g_{2a}, g_{2b} and g_3 are all inequalities. It should be noted that there is a g_1 equation for each edge feature that is to be resolved, while for the depth-of-field and field-of-view relationships, there is a single g_{2a}, g_{2b} and g_3 for all features.

While the constraints address the admissibility of the computed solution, the optimization function on the other hand is constructed so as to characterize the "quality" of the computed solution. The measure used to assess the goodness of a solution with respect to the resolution, field-of-view and depth-of-field constraints, is the value of the constraint relationships g_i, $i = 1,2a,2b,3$ themselves. This is appropriate since a large positive value of g_i indicates that the constraint is satisfied comfortably, a small positive value indicates marginal satisfaction, while inadmissible solutions give rise to negative values. Similarly for the visibility constraint, a measure of this type needs to be also formulated. For this purpose, the minimum distance, d_v, from the viewpoint to the polyhedron describing the visibility region seems suitable. More specifically, the distance measure is chosen to be: $g_4 = \pm d_v$. - where $+d_v$ or $-d_v$ depending on whether the point is inside or outside the visibility volume respectively. The optimization function is taken to be a weighted sum of the above component criteria, each of which characterizes the quality of the solution with respect to each associated requirement separately. Thus the optimization function can be written as:

$$f = max\ \alpha_1 g_1 + \alpha_{2a} g_{2a} + \alpha_{2b} g_{2b} + \alpha_{3b} g_3 + \alpha_4 g_4$$

subject to $g_i \geq 0$, $i = 1,2a,2b,3,4$ and $g_5 = 0$, where α_i are the weights. These weights are currently chosen so that the contribution of each constraint to the objective function is of the same order of magnitude, and in this way avoid having a subset of the constraints dictate the optimization. In future work we shall investigate other weight settings as well. Given the above formulation, the optimization starts with an initial point in the domain of possible generalized viewpoints and then generates a generalized viewpoint that is globally admissible and locally optimal as described by the optimization function. In other words, all constraints are satisfied with the largest margin in a neighborhood of the initial point.

5 EXPERIMENTS

As part of the MVP system, we have implemented the machine vision planning algorithms that were discussed in the previous section. In the experiments, we demonstrate the effectiveness of this approach using a robot vision system that plans its pose and the lens settings of its camera according to these techniques.

A CAD model of the object that is used in the camera placement experiments is shown in in Figure 3. The features to be observed are the two edges of the enclosed cube shown in Figure 3.

The domain of admissible camera locations is initially limited to the region in three-dimensional space from where the edge features to be observed are *visible*. The visibility region for each edge is computed separately by the method discussed in section 3.1 and then these two regions are intersected in order to determine the region in space from where both edges are simultaneously visible. The two edge visibility regions when intersected result in the region shown in Figure 1, that is the space from where both edges are simultaneously visible.

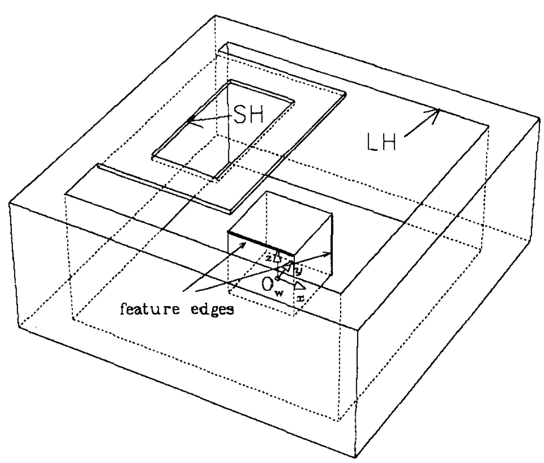

Figure 3. CAD model of the object.

Viewpoints chosen from this visibility region must also satisfy the other constraints in order to be globally admissible. For this experiment the resolution specification is taken to be 1 frame buffer pixel spacing per $l = 0.1$

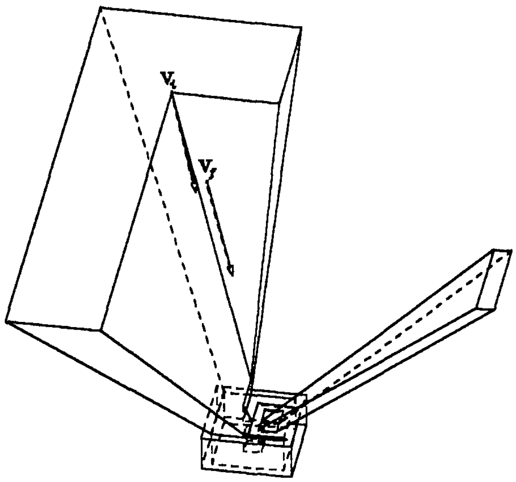

Figure 1. The visibility region for both edges.

inches.. Since the diagonal direction in the sensor plane corresponds to the worst case, that is, it yields the minimum resolution [3], we compute the sensor plane spacing in the diagonal direction corresponding to 1 frame buffer pixel spacing: $w = \sqrt{(23 \times .70642)^2 + 13.5^2}$ $= 21.12$ *microns* where the spacing between sensor elements in the horizontal and vertical directions are respectively 23 and 13.5 microns and the ratio of the sensor element spacing in the horizontal direction to the picture element spacing after sampling by the image acquisition

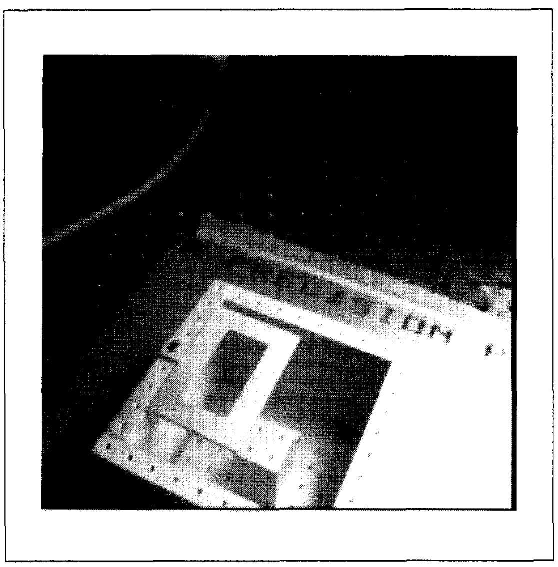

Figure 2. The camera view of the features from the computed viewpoint.

hardware was found to be 0.70642 from calibration [6]. This horizontal scale factor relates the sensor element spacing to the pixel spacing in the frame buffer. For the field-of-view constraint, the minimum sensor plane dimension is $I_{\min} = min\,(479 \times 13.5\,,\,383 \times 21) = 479 \times 13.5$ $\simeq 6.5\,mm$ where 480 and 384 are the number of sensor elements in the vertical and horizontal directions respectively for the Javelin CCD camera at hand. The diameter of the circle of confusion for the depth-of-field constraint is taken to be the minimum of the horizontal and vertical sensor element spacings, that is $c = 13.5$ *microns*.

With this information the optimization constraints and objective function are constructed as given by g_i, $i = 1,2a,2b,3,4,5$ and f. In this experiment, values of the lens aperture a and the intrinsic focal length f were chosen a priori ($f = 12.5mm$ and $a = f/16$) and thus, values for the remaining imaging space parameters $x, y, z, \vec{v}$ and d were computed. The contribution of the visibility constraint to the objective function as given by g_4 (see section 4) has not yet been implemented. However, the point-in-polyhedron classification is incorporated and consequently satisfaction of the visibility requirement can be determined. Using this classification, the final point as determined by the optimization is classified with respect to the visibility volume. If this point lies inside or on the visibility volume, then it is a globally admissible generalized viewpoint that is locally optimal with respect to the resolution, depth-of-field and field-of-view constraints, and thus, it is the viewpoint of choice. However, if the viewpoint lies outside the visibility volume, then the intermediate points that are generated by the optimization are checked for global admissibility, that is, satisfaction of all the constraints including visibility. From amongst these, the one with the largest value of the objective function is chosen. However, if no such globally admissible intermediate points exist, then the optimization is performed again with a different initial viewpoint. The values of the weights α_i in the optimization function were taken to be: $\alpha_1 = 10^3$, $\alpha_{2a} = \alpha_{2b} = 10^{-2}$ and $\alpha_3 = 10^{-1}$, when distances are expressed in millimeters. The optimization is performed using the IMSL non-linear constrained optimization routine NCONF.

Both the initial and the computed camera viewpoints are listed in Table 1 and are shown in Figure 1 as points V_i and V_f respectively along with their associated viewing vectors. It can be seen from Figure 1 that the initial guess viewpoint for the optimization is chosen to lie on an edge of the visibility volume with a viewing vector in the direction from the viewpoint to the center of the sphere that circumscribes the features to be observed (see section 3.4). The viewpoint determined by the optimization, V_f, is classified with respect to the visibility volume and is determined to lie inside the visibility region as can be seen in Figure 1. Thus, this viewpoint is both globally admissible and locally optimal.

Table 1. The initial and final generalized viewpoints									
	x	y	z	v(1)	v(2)	v(3)	f	d	a
V_i	165.1 mm	-279.4 mm	368.3 mm	-0.336	0.569	-0.750	12.5 mm	1.2f	f/16
V_f	107.58 mm	-231.43 mm	258.68 mm	-0.389	0.694	-0.608	12.5 mm	13.07 mm	f/16

Having determined from optimization the position vector of the front nodal point of the lens, the optical axis orientation and the image plane to back nodal point distance, the camera can be placed and focused accordingly. Details regarding the camera placement computations can be found in [10]. On the other hand, the image plane to back nodal point distance, d, that has been computed by the optimization needs to be realized in the camera-lens setup. From calibration of the lens [15], the image plane to back nodal point distances d can be determined for the $f = 12.5\ mm$ lens at its various focus settings. The value of this distance for the computed viewpoint is $d = 13.07 mm$, as seen in Table 1, and is found to lie outside the the limits of the focusing capability of the lens, as determined by calibration. As a result a lens extension is used in order to give the necessary lens to image distance d. The manipulator is placed at viewpoint V_f and oriented according to the computed viewing vector.

The scene of the object taken from the computed viewpoint V_f is shown in Figure 2. Satisfaction of the visibility and field-of-view constraints can be readily verified. The former however is satisfied only marginally since the horizontal edge is very close to being occluded by the overhang. This emphasizes the importance of including the contribution to the objective function from the visibility constraint as discussed in section 4 and thus avoid viewpoints close to the visibility volume boundary. Furthermore, the resolution requirement is verified by measuring the feature magnification in the image. The 1 inch horizontal and vertical edges are imaged as 45 and 39 frame buffer pixels, thus comfortably satisfying the 1 frame buffer pixel per 0.1 inch requirement. Finally, the depth-of-field requirement can also be qualitatively verified in the image. A quantitative measure of focus will also be employed in future work for a more accurate evaluation of the quality of focus.

6 CONCLUSION

We presented a method to determine optimal sensor placement and optical settings of a camera, so that given visibility, field-of-view, depth-of-field and resolution requirements are simultaneously satisfied with margin for chosen object features. This is achieved by merging the analytical loci of admissible camera poses and settings that we had determined for each requirement in previous work. The problem is posed in an optimization setting and a viewpoint is sought that is both globally admissible and central to the feasibility domain. This approach provides advantages over the currently employed techniques in which sensor configurations are generated and then tested for satisfaction of the task requirements. The results are valid for a general three-dimensional viewing configuration and were demonstrated using a robot vision system.

This research can be extended to include illumination parameters (e.g. illuminator pose) and investigate how they can be planned in a given task so that features of interest are again robustly *detectable* in the resulting image.

The results discussed in this paper are useful for automating the vision system design process, as well as for programming the vision system itself. In addition, such planning techniques can also automate robot imaging systems that reconfigure themselves in an intelligent manner in order to optimize imaging quality.

7 BIBLIOGRAPHY

[1] Cameron, A., and Durrant-Whyte, H., A Bayesian approach to optimal sensor placement, Oxford, UK: Department of Engineering Science, University of Oxford, OUEL 1759/89, 1989.

[2] Cowan, C. K., and Bergman, A., "Determining the camera and light source location for a visual task," *Proceedings 1989 IEEE International Conference on Robotics and Automation*, pp. 509-14, 1989.

[3] Cowan, C. K., and Kovesi, P. D., "Automatic sensor placement from vision task requirements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 3, pp. 407-16, May 1988.

[4] Hager, G. D., *Active reduction of uncertainty in multi-sensor systems*, PhD thesis, University of Pennsylvania, 1988.

[5] Krotkov, K., *Exploratory visual sensing with an agile camera*, PhD thesis, University of Pennsylvania, 1987.

[6] Lenz, R. K., and Tsai, R. Y., "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 5, pp. 713-20, Sept. 1988.

[7] Niepold, R., Sakane, S., Sato, T., and Shirai, Y., "Vision sensor set-up planning for a hand-eye system using environmental model," *Proc. Soc. Instrum. Control Eng. Japan*, pp. 1037-1040, Hiroshima, Japan, July 1987.

[8] Sakane, S., Ishii, M., and Kakikura, M., "Occlusion avoidance of visual sensors based on a hand eye action simulator system: HEAVEN," *Adv. Robot.*, vol. 2, no. 2, pp. 149-65, 1987.

[9] Shafer, S. A., Automation and calibration for robot vision systems, Pittsburgh, PA: Carnegie Mellon University, CMU-CS-88-147, May 1988.

[10] Tarabanis, K., and Tsai, R. Y., Camera placement planning avoiding occlusion: Test results using a robotic hand/eye system., New York, NY: Computer Science Department, Columbia University, CUCS-501-89, October 1989.

[11] Tarabanis, K., and Tsai, R. Y., "Viewpoint planning: the visibility constraint," *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, May 23-26 1989.

[12] Tarabanis, K., and Tsai, R. Y., Computing viewpoints, Yorktown Heights, NY: IBM T.J. Watson Research Center, RC 16312, November 1990.

[13] Tarabanis, K., Tsai, R. Y., and Allen, P. K., Automated sensor planning and modeling for robotic vision tasks, Yorktown Heights, NY: IBM T.J. Watson Research Center, RC 16178, October 1990.

[14] Tarabanis, K., Tsai, R. Y., and Allen, P. K., "Satisfying the Resolution Constraint in the "MVP" Machine Vision Planning System," *13th IASTED International Symposium Robotics and Manufacturing, also in Proc. 1990 DARPA Image Understanding Workshop*, 1990.

[15] Tsai, R. Y., "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses," *IEEE Journal Robotics Automation*, vol. RA-3, no. 4, August 1987.

[16] Tsai, R. Y., and Tarabanis, K., "Model-based planning of sensor placement and optical settings," *Proc. Sensor Fusion II: Human and Machine Strategies*, Philadelphia, Pennsylvania, November 6-9 1989.

[17] Tsai, R. Y., and Tarabanis, K., "Occlusion-free sensor placement planning," in Freeman, H., editor, *Machine Vision for Three-Dimensional Scenes*, San Diego, California: Academic Press, 1990.

[18] Yi, S., Haralick, R. M., and Shapiro, L. G., "Automatic sensor and light source positioning for machine vision," *Proceedings of the 10th International Conference on Pattern Recognition*, pp. 55-59, Atlantic City, NJ, June 16-21 1990.