Optimal Stereo Mast Configuration for Mobile Robots

Wesley H. Huang Eric P. Krotkov
Robotics Institute
Carnegie Mellon University

Abstract

Planning the placement of stereo cameras on a mobile robot is often a balancing act between the quality of stereo data, the reliability of stereo matching, and ensuring that the cameras see enough of the robot's environment for navigation. In this paper, we present a nonlinear programming formulation to determine the optimal parameters for two stereo pairs of cameras and a camera mast for a mobile robot. We have applied this method to the design of a stereo camera mast for *Nomad*, a prototype rover currently under construction at Carnegie Mellon University, and present the results of designing that mast.

1 Introduction

Mobile robots need to sense their environment in order to navigate in the world. One rich source of information about the world is visual information. From a pair of cameras, we can get three dimensional information about the world around a robot.

The choice of a stereo vision algorithm affects processing time and the quality of the resulting data, but it is the placement and configuration of the cameras that has the greatest effect on the overall utility of stereo data. Camera placement and configuration must balance a number of sometimes conflicting criteria: there are physical constraints upon where the cameras can be mounted on the robot, the cameras must be able to see as much of the area around the robot as possible, the robot should not obstruct the camera's view, there must be enough resolution, we must be able to process the information fast enough relative to the speed of the robot, etc.

Often the choices for camera parameters and placement are made hastily, guided by human intuition and experience. As we build robots for situations that increasingly demand greater performance, such as sending robots into hazardous environments, space, or other planets, we would like to get the most out of our visual sensing.

This paper describes the formulation of a dynamic program to optimize the parameters for a stereo camera mast for a mobile robot. We assume that the location on the robot for this camera mast has been fixed, and we are to determine the optimal parameters for the height of the mast, the stereo baseline, the angular orientation of the cameras, and the focal length of the cameras for a horizontal stereo pair.

The criteria for optimization will in general depend upon the tasks which the robot is required to perform. Here we have primarily considered a robot that must map and navigate in an expansive environment. Consequently, the measure of optimality we use is the area of stereo coverage — a mapping and navigating robot must see as much of its environment as possible. We discuss the objective function for optimization in section 5.

First, however, we describe related work, present a formal problem statement, and formulate the constraints. We conclude by describing the application of this method to configuring the stereo mast for *Nomad*, a rover currently under construction at Carnegie Mellon.

2 Related Work

Our work is most closely related to the MVP system developed by Tarabanis *et al.* [10]. The MVP system plans single camera views of static scenes by creating a nonlinear program. The constraints ensure that desired features in a scene are visible, will have sufficient resolution, are all in focus, and are within the field of view of the camera. The objective function measures the robustness of the viewpoint (i.e. how "easily" it satisfies the constraints).

Some early work in planning camera placement for inspection or light placement for photometric stereo include Cowan and Kovesi [1] and Sakane [7]. A survey of camera placement planning appears in Tarabanis *et al.* [9].

In the field of visual servoing, there has been work by Nelson and Khosla [5] and Sharma and Hutchinson [8] in developing methods for camera placement such that the best possible visual feedback is provided. Nelson and Khosla address the case of stereo cameras as well as a single camera.

There has also been work in the field of teleoperation to determine the effect of camera placement on effectiveness. Miller and McGovern [4] and Pisanich *et al.* [6] performed experiments to test the effect of camera placement. The former studied teleoperation of a jeep along a dirt road; the latter studied heuristic camera placement techniques in conjunction with a robotic manipulation task.

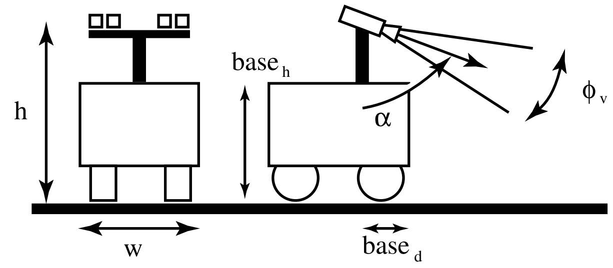

Figure 1: Configuration of the stereo mast. We assume that the location of the mast, i.e. the distance from the front of the robot ($base_d$) and the height of its base ($base_h$), is given. Our optimization will determine the height of the camera mast (actually the height of the cameras above the ground, h), the elevation of the cameras (α), and the focal length of the lens (f). The focal length will determine the vertical field of view (ϕ_v).

This paper studies camera placement in the context of using multiple pairs of stereo cameras on a mobile robot for obstacle detection in an expansive unstructured environment. This results in a very different set of constraints to guide camera placement. Camera parameters and placement are optimized using a nonlinear program. The constraints presented in this paper are based on work by Krotkov and Hebert [2].

3 Problem Statement

We assume that we are given a mobile robot and that the stereo mast will consist of a vertical strut with a horizontal crosspiece upon which the cameras will be mounted. The location of the stereo mast is given; see figure 1.

In our experience, two pairs of stereo cameras are required in order to give sufficient width of coverage; this is the case we will develop here, though our procedure can be followed with any number of pairs of cameras.

As illustrated in figure 2, we require each stereo camera pair to have the same baseline; one camera of each pair is on the left of the mast, the other on the right. The rightmost cameras on each side of the mast point towards the right, the leftmost to the left.

The independent variables that we optimize are:

- h the height of the cameras
- b the baseline of the stereo pairs
- f the focal length of the cameras
- α the elevation of the cameras (from vertical) of the cameras
- β the azimuthal rotation of the cameras (from straight ahead)

3.1 Physical Parameters

We assume a number of parameters about the robot are given. We divide these parameters into two categories, physical parameters (dimensions of the robot, etc.) and dependent variables (variables whose values are functions of the independent variables).

The physical parameters are:

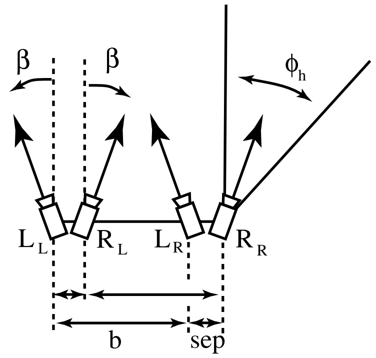

Figure 2: Configuration of the cameras on the stereo mast. The left stereo pair is marked L_L and L_R for the left and right camera, respectively. The right stereo pair is marked R_L and R_R. The separation between the two right cameras and two left cameras (sep) is given; the optimization will determine the azimuthal rotation of the cameras (β), the baseline of the cameras (b), and the focal length of the lens (f). The focal length will determine the horizontal field of view (ϕ_h).

- w the width of the robot
- s_v the vertical dimension of the CCD element
- s_h the horizontal dimension of the CCD element
- $base_h$ the height (above ground) of the base of the mast
- $base_d$ the distance from the front of the robot to the base of the mast
- sep the separation between adjacent cameras (between each of the "left" cameras and between each of the "right" cameras)

3.2 Dependent Variables

The dependent variables are:

- ϕ_h the horizontal field of view of the cameras
- ϕ_v the vertical field of view of the cameras
- d_l the lookahead distance

The horizontal and vertical fields of view are based upon the focal length of the lens and the size of the CCD element. We have:

$$\phi_h = 2\tan^{-1}\frac{s_h}{2f} \tag{1}$$

$$\phi_v = 2\tan^{-1}\frac{s_v}{2f} \tag{2}$$

The lookahead distance is the minimum distance at which the robot must be able to detect an obstacle and come to a stop (from maximum velocity) before hitting that obstacle. Although technically dependent upon camera resolution and the minimum obstacle size (since these factors affect the processing time of object recognition programs), the lookahead distance is dominated by detection time (t_{stop}), the maximum velocity of the robot (v_{max}), and the braking distance of the robot (d_{brake}). We can write:

$$d_l = t_{stop}v_{max} + d_{brake} \tag{3}$$

Practically speaking, however, the lookahead distance may be determined by the situation, or the minimum lookahead distance (considering hardware constraints) may determine the maximum velocity of the robot.

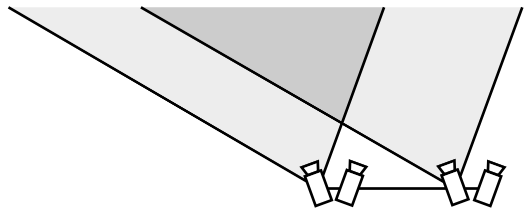

Figure 3: Stereo coverage. A pair of cameras can provide stereo coverage in the "cone" shaped intersection of their fields of view.

4 Constraints

4.1 Camera Orientation

Our first set of constraints simply requires the cameras to point in the correct direction, ensuring the validity of later calculations.

First of all, since we are interested in sensing the terrain in front of the robot the cameras must look forwards and downwards:

$$\alpha > 0 \tag{4}$$

$$\alpha - \frac{\phi_v}{2} < \frac{\pi}{2} \tag{5}$$

In the horizontal plane, we want to ensure that the left pair points to the left (and vice versa) and that the cameras are pointing forwards:

$$\beta > 0 \tag{6}$$

$$\beta + \frac{\phi_h}{2} < \frac{\pi}{2} \tag{7}$$

Finally, we must have a positive baseline:

$$b > 0 \tag{8}$$

4.2 Distortion

The minimum focal length in order to avoid significant barrel distortion on a $\frac{2}{3}$ inch CCD camera was empirically determined to be 8 mm:

$$f > 0.008 \tag{9}$$

We also add a maximum focal length constraint to keep the focal length at a reasonable value and to help the nonlinear program solver stay on track:

$$f < 0.050 \tag{10}$$

4.3 Width of Coverage

We want to see 3 times the width of the robot at the lookahead distance — this allows for the robot to steer around any obstacle it can detect; the factor of 3 generously allows for control and sensing tolerances.

There are two parts to this constraint. First, the fields of view of the two stereo pairs must overlap. Second, the width of coverage must be wide enough. Figure 3 illustrates the stereo coverage of a single pair of cameras; figure 4 shows the overlapping stereo coverage from two stereo pairs.

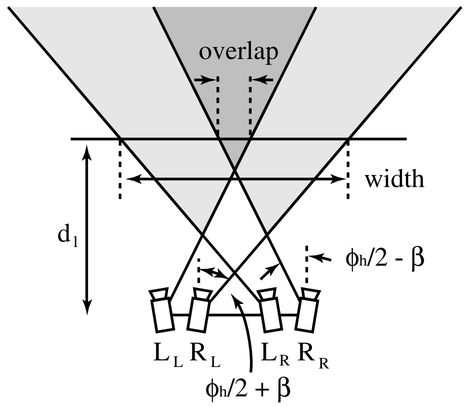

Figure 4: Width of coverage. Two pairs of cameras provide a greater area of stereo coverage. We must have a certain width of stereo coverage at the lookahead distance; the stereo coverage must also overlap at the lookahead distance.

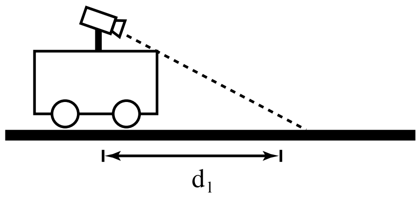

Figure 5: Clearance. If the camera height is too low, the cameras cannot see the ground at the lookahead distance.

The left edge of camera R_R (see figure 4) and the right edge of camera L_L determine the overlap of the fields of view. We want to ensure that the stereo fields of view overlap at least d_o at the lookahead distance. We have

$$2d_l \tan(\frac{\phi_h}{2} - \beta) - b - sep \geq d_o \tag{11}$$

The width of coverage (wd) is determined by the inside edges of the inside cameras (right edge of R_L and left edge of L_R). See figure 4. We have

$$wd = 2d_l \tan(\beta + \frac{\phi_h}{2}) - b + sep \geq 3w \tag{12}$$

4.4 Clearance

Simply stated, the cameras must be able to see the ground at the lookahead distance without being obstructed by parts of the robot. This is a constraint on the height of the mast, not the inclination of the cameras (which will be handled separately). As shown in figure 5, if the mast is too low, the cameras will not be able to see to the lookahead distance regardless of the camera inclination.

This constraint could take many forms. However, given the shape of the robot and given the lookahead distance, we can compute a simple minimum bound on the height (h_{min}).

$$h \geq h_{min} \tag{13}$$

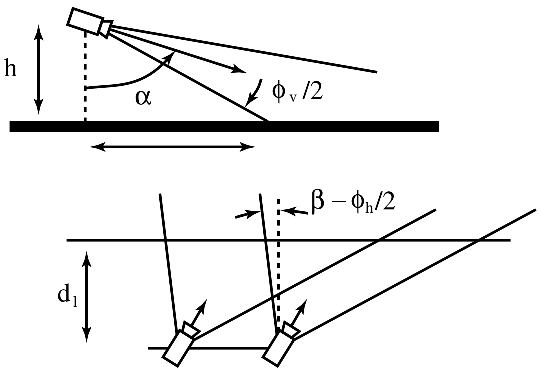

Figure 6: Lookahead. In order to see the ground at the lookahead distance, the cameras must point downwards far enough. However, the further the cameras are rotated outwards, the lower the elevation of the cameras must be.

4.5 Lookahead

Assuming that the mast is high enough to satisfy the clearance constraint, we must make sure that the cameras are tilted so that they can see objects at the lookahead distance. Primarily, this is a constraint on the camera inclination, α. However, as figure 6 illustrates, it will also involve β. This constraint ensures that the closest point seen by the cameras is at least as close as the lookahead distance.

$$h \tan(\alpha - \frac{\phi_v}{2}) \leq \frac{d_l}{\cos(\beta - \frac{\phi_h}{2})} \tag{14}$$

4.6 Error

We wish to have an acceptable depth error due to a 1 pixel error in disparity. Arguably, this might have been a suitable quantity for an objective function to minimize. However, beyond a certain error margin, decreasing the error further does not provide useful information and actually makes stereo matching harder.

Depth from stereo is given by $z = \frac{bf}{disp}$ where $disp$ is the disparity. We must pick a target range in order to compute this error; we choose this distance to be the lookahead distance.

The disparity at the lookahead distance is $disp = \frac{bf}{d_l}$. If the disparity at the lookahead distance is off by one pixel (we assume it is one pixel less than it should be, for this produces the greater distance error), then the calculated distance will be

$$z = \frac{bf}{disp - pixsize}$$

The error is then

$$z - d_l = \frac{d_l}{\frac{bf}{(pixsize)(d_l)} - 1} \leq e_{max} \tag{15}$$

where we assume that $pixsize = \frac{s_h}{512}$ is the size of a pixel (or more accurately, the spacing between pixels).

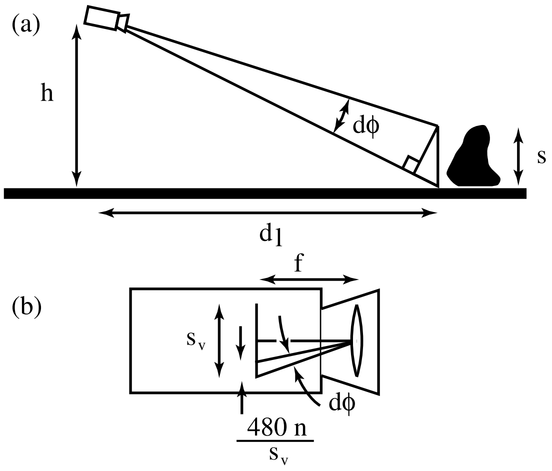

Figure 7: Resolution. The number of pixels on an object of height s at the lookahead distance can be determined by comparing similar triangles formed by the camera and the object and by the camera and the object's image.

4.7 Resolution

We want to ensure that we have enough pixels on an obstacle so that it can be detected. We assume that we want to be able to detect an obstacle of height s at the lookahead distance and require that the object occupy at least n_{min} pixels in the image. See figure 7a. Using similar triangles, we find that

$$\frac{d\phi\sqrt{d_l^2 + h^2}}{s} = \frac{d_l}{\sqrt{d_l^2 + h^2}}$$

So

$$d\phi = \frac{s\, d_l}{d_l^2 + h^2}$$

Assuming that this is imaged at the edge of the CCD, and again using similar triangles (see figures 7 a and b), we find that

$$\frac{d\phi\sqrt{f^2 + \frac{s_v}{2}}}{\frac{n}{480}s_v} = \frac{f}{\sqrt{f^2 + (\frac{s_v}{2})^2}}$$

assuming that there are 480 pixels in the vertical direction. Solving for the number of pixels, n, we get

$$n = 480\frac{f^2 + (\frac{s_v}{2})^2}{f s_v}\frac{s d_l}{d_l^2 + h^2} \geq n_{min} \tag{16}$$

5 Objective Function

A robot that maps and navigates in an expansive unstructured environment needs as much information as possible about its environment. This can be accomplished by maximizing the area of stereo coverage.

For cameras that are pointing downward, a stereo pair sees a trapezoidal region on the ground (see figure 8). Aside from the complexity of calculating the total area of two intersecting trapezoids, we encounter the problem that

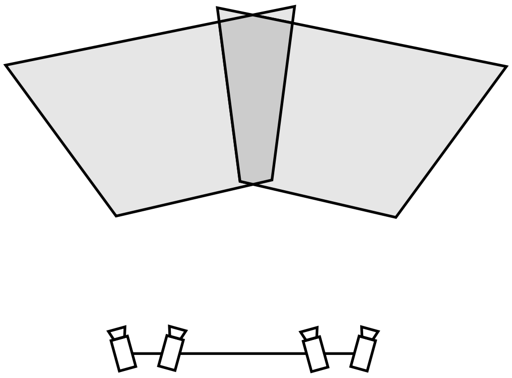

Figure 8: Stereo coverage. The true area of stereo coverage is not simple to compute and is not well defined when the cameras see above the horizon.

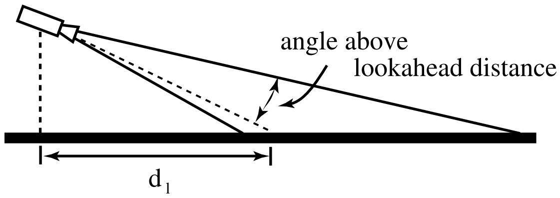

Figure 9: Stereo coverage approximation. An approximate measure of stereo coverage is the product of the width of coverage at the lookahead distance and the angle of coverage above the lookahead distance.

this area approaches infinity as the top of the frame approaches the horizon. Furthermore, it is not clear what to do when the top of the image sees above the horizon.

For these reasons, we have adopted a slightly modified version of this idea. Instead of maximizing the coverage area, we maximize the product of the width of coverage at the lookahead distance and the angle that we can see above the lookahead distance; see figure 9. In some sense, this is an estimate of the solid angle of coverage.

Our objective function is:

$$wd(\alpha + \frac{\phi_v}{2} - a_{min}) \tag{17}$$

where wd is as defined in equation 12 and $a_{min} = \tan^{-1}\frac{d_l}{h}$.

6 Application to Design of a Rover

Nomad is a mobile robot being developed at Carnegie Mellon University in anticipation of a future lunar rover mission; profiles of the robot are shown in figure 10. We applied the procedure described in this paper to design the stereo camera mast for this robot.

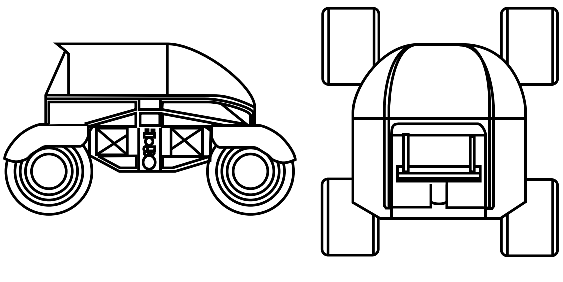

Figure 10: Profiles of the *Nomad* rover.

w	2.5 m	width of robot
s_v	0.0066 m	height of CCD
s_h	0.0088 m	width of CCD
$base_h$	1.50 m	height of mast base
$base_d$	1.18 m	distance from mast to front
sep	0.05 m	camera separation
d_o	0.50 m	stereo overlap
h_{min}	1.70 m	minimum height
e_{max}	0.05 m	max 1 pixel disp. error
s	0.30 m	minimum obstacle size
n_{min}	12	min pix on min obstacle

Table 1: Physical parameters and other constants for the design of the *Nomad* stereo camera mast.

6.1 Constraints for *Nomad*

The location of the stereo camera mast and some limitations on its size were given because the mast must fold back inside the robot for compactness (in order to fit into the payload section of a rocket). The mast could be no taller or wider than 1 meter without requiring some sort of telescoping mechanism. In order to reduce mechanical complexity, it was strongly desired to configure the mast within these limits.

Nomad is a very wide robot; fully expanded, it is approximately 2.5 meters square and 1.5 meters high (without the stereo mast). We had some concern as to whether two stereo pairs could produce coverage that was 3 times the width of the robot at the lookahead distance. However, hardware constraints strongly discouraged the use of three stereo pairs.

The physical parameters and other parameters for *Nomad* are given in table 1. We tried several different values for the lookahead distance.

6.2 The Nonlinear Program Solver

We used GINO [3], a nonlinear programming package, to solve the nonlinear program. GINO uses a version of the generalized reduced gradient algorithm for nonlinear optimization.

The entire GINO program for this example can be found on the World Wide Web at:
`http://www.cs.cmu.edu/˜mongoose/research/nomad.html`.

6.3 Results

The results of three runs of the nonlinear program solver are shown in table 2. The first run was without any con-

	Run		
Variable	1	2	3
h	3.71 m	2.50 m	2.50 m
b	0.59 m	0.59 m	0.53 m
f	8 mm	8 mm	8 mm
α	67.5°	78.5°	77.0°
β	20.0°	20.0°	20.0°
objective	6.206	6.203	5.893

Table 2: Results of three runs of the nonlinear program solver. The first run was without any height constraint; the second included a height constraint. The third run used a reduced lookahead distance.

straint on the height and produced a height that was higher than acceptable. The second run was done with a height limit of 2.5 meters, and not surprisingly, the resulting solution used the full height.

The first two runs used a lookahead distance of 2.5 meters. We tried several runs in which we gradually reduced this distance until there was no longer a feasible solution. The third run uses the minimum lookahead distance (of 2.3 meters) which still has a feasible solution.

GINO also provides some "price" information — an estimate of the improvement in the objective function if each of the constraints is relaxed. With this information, we can gain some insight into which factors are most important in this problem. The "highest priced constraints" were, not surprisingly, the focal length and the size of the CCD; increasing the CCD size or decreasing the focal length would increase the field of view of the cameras. The constraint on error due to a one pixel disparity mismatch had the next highest price, followed by the amount of stereo coverage overlap required at the lookahead distance. Neither the maximum height constraint nor the resolution constraint had a high price.

7 Conclusions

In this paper, we formulated a nonlinear program to determine the optimal configuration for a multi-pair stereo camera mast for a mobile robot. We require that any valid configuration be able to see obstacles on the ground at the lookahead distance; there must be sufficient resolution in order to detect an obstacle; the robot should see three times the width of the robot at the lookahead distance; and there must be an acceptable depth error due to a one pixel error in disparity. We optimize based upon an approximation to the area of coverage of the cameras.

We applied this technique to the design of *Nomad*, a prototype rover currently under construction at Carnegie Mellon. We found that this method produced reasonable results and were pleasantly surprised that the resulting solution called for a stereo baseline that was shorter than what had been picked for previous robots (0.59 meters instead of 1 meter). As this is a valid solution, this baseline satisfies error constraints and should make stereo matching easier than it was with the longer baseline.

Through "price" information returned by our nonlinear program solver, we learned that satisfying the maximum height constraint and getting sufficient resolution on obstacles are not difficult. The stereo coverage can be most dramatically improved by increasing the acceptable depth error from disparity mismatch, by decreasing the amount of stereo coverage overlap required at the lookahead distance, and by increasing the field of view of the cameras.

Acknowledgements

We would like to thank Martial Hebert, Al Kelly, and Mark Maimone for engaging in several discussions of this work. This work was supported by NASA through grant NAGW 1175 and through the Graduate Student Researchers Program.

References

[1] C. Cowan and P. Kovesi. Automatic sensor placement from vision task requirements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:407–16, 1988.

[2] E. Krotkov and M. Hebert. Design considerations on the stereo system for the 1994 lunar rover demonstration. Unpublished report, February 1994.

[3] J. Liebman, L. Lasdon, L. Schrage, and A. Waren. *Modeling and Optimization with GINO*. The Scientific Press, 1986.

[4] D. P. Miller and D. E. McGovern. A laboratory-simulation approach to the evaluation of vision systems for teleoperated vehicles. In *International Symposium on Teleoperation and Control*, pages 85–94, 1988.

[5] B. Nelson and P. Khosla. Vision resolvability for visually servoed manipulation. *Journal of Robotic Systems*, 13(2):75–93, February 1996.

[6] G. M. Pisanich, M. P. Prevost, and S. B. Hall. Evaluating the impact of camera placement on teleoperator efficiency. In *International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, pages 1:629–37, 1988.

[7] S. Sakane and T. Sato. Automatic planning of light source and camera placement for an active photometric stereo system. In *IEEE International Conference on Robotics and Automation*, pages 2:1080–7, 1991.

[8] R. Sharma and S. Hutchinson. Optimizing hand/eye configuration for visual-servo systems. In *IEEE International Conference on Robotics and Automation*, pages 1:172–7, 1995.

[9] K. A. Tarabanis, P. K. Allen, and R. Y. Tsai. A survey of sensor planning in computer vision. *IEEE Transactions on Robotics and Automation*, 11(1):86–104, February 1995.

[10] K. A. Tarabanis, R. Y. Tsai, and P. K. Allen. The mvp sensor planning system for robotic vision tasks. *IEEE Transactions on Robotics and Automation*, 11(1):72–85, February 1995.