OPTIMAL CAMERA PLACEMENT FOR AN ACTIVE-VISION SYSTEM

X. He†
PhD. Candidate

B. Benhabib†*
Associate Professor

K. C. Smith†*‡
Professor

R. Safaee – Rad†
Research Director#

\# Computer Integrated Manufacturing Laboratory
Department of Mechanical Engineering
University of Toronto
Toronto, Ontario
Canada M5S 1A4

† — Dept. of Mechanical Engineering, * — Dept. of Electrical Engineering, ‡ — Dept. of Computer Science

Abstract — **Image-analysis in an industrial environment can be simplified by pre-marking the objects using artificial *markers* and active vision to acquire pre-specified views of these objects. In this context, a problem which should be addressed is the determination of the initial locations of the cameras (including one mobile camera) such that marker(s) can be efficiently detected. A method for determining the optimal number of cameras and their best locations is presented in this paper, given the objects' geometries and the positions of markers on them. Such an optimal camera-arrangement would yield maximum detectability of markers and also satisfy a user given threshold on the probability of detecting at least one marker. Analytic-geometry, probability theory and CAD databases are used in solving this problem.**

I. INTRODUCTION

In the context of a new active-vision system under development in the Computer Integrated Manufacturing Laboratory at the University of Toronto, special features — *markers* — placed on an object should be detected to facilitate the 3D-recognition process. In this active-vision system, an object is modeled using only a small number of its 2D distinct perspective views [1]. These are referred to as *standard-views*, with each view having a corresponding *standard-view-axis*. For successful recognition, the input image of an object must be one of its standard perspective views. A standard-view is acquired by aligning the optical axis of a mobile camera with one of the standard-view-axes of the object. The matching process is then performed between the acquired 2D standard-view of the object under consideration and the library of 2D standard-views of a set of objects.

To enable the vision system to acquire standard-views, standard-view-axes are pre-defined. This can be accomplished by defining a local surface normal for each distinct view of an object. These surface normals can be defined by adding nonfunctional markers to an object, preferably during the manufacturing process.

The successful determination of a marker's normal and position is an important issue in the context of the proposed vision system. One of the problems addressed is the optimal arrangement of the initial camera locations, of which one is the mobile camera. The primary role of the other cameras employed would be only the detection of markers.

The problem of the optimal placement of cameras/sensors has been studied by various researchers. Examples include: the determination of the optimal sensor locations, based on the acquired partial knowledge of the viewed object, for performing recognition and localization operations [2], and for performing the information fusion process [3]; the determination of the optimal placement and optical settings of the vision sensors for satisfying particular image constraints [4]; and, the determination of the optimal sensor and light source positions for edge-detection [5]. In our research, we address a similar optimization problem though in a different context, that is, the determination of the optimal number and initial locations of the cameras for the detection of features of a known pattern (the markers) on the objects.

II. PROBLEM STATEMENT

Two criteria are considered for the optimization problem: One is the minimization of the number of cameras, since it would not be desirable to have a large number of cameras. The other is the maximization of the detectability of the markers — geometric visibility and optical detectability. In this paper, however,

ISSN# 0-7803-0233-8/91 $1.00©1991 IEEE

those factors such as illumination, image resolution, etc., which would affect the optical detectability, are not considered.

The problem is considered under the conditions that: the number and locations of markers on the objects are pre-defined, and the probability of detecting at least one marker on an object viewed by the camera(s) should not be less than a user chosen threshold.

The optimization problem can then be formulated as: Given a set of pre-marked objects, determine the optimal number and positions of the cameras such that: the number of cameras (N_{cm}) is minimized, the detectability of the markers (J_{dt}) is maximized, and the probability of detecting one marker is not less than a user chosen threshold. This is actually a two level optimization, and can be expressed as:

$$
\begin{array}{rl}
min. & N_{cm} \\
then\ max. & J_{dt}(\gamma),\ for\ current\ N_{cm} \\
subject\ to & \left\{ \begin{array}{l} \gamma \in \Gamma \\ N_{cm} \leq n^* \end{array} \right.
\end{array} \tag{1}
$$

where γ denotes a variable (or a set of variables) defining the locations of the cameras, Γ denotes the feasible range of γ determined by the given threshold on the probability of detecting a marker, and n^* is a user defined upper limit on N_{cm}.

III. SOLUTION METHOD

3.1 Assumptions

The main assumptions made in our proposed solution method are as follows: (1) Each object is viewed separately (isolated objects), and the viewed object is located at the center of the object-space; (2) The cameras are located on the surface of a pseudo-sphere centered at the center of the object-space, while the cameras' optical axes point to the center of the sphere; (3) The field-of-views of all the cameras are large enough to cover the viewed object by using proper lenses and/or proper camera-object distances; and, (4) There exist only a finite number of stable-orientations for each object, with known probable occurrence frequencies.

3.2 Definitions of camera-locations and a marker's visible-space

The camera locations, as mentioned before, are defined to be on the surface of a pseudo-sphere, referred to as the *location-sphere*. It is assumed that when an object appears in the scene at one of its stable-orientations, the rotation of the object about the normal of its "supporting-plane" is not known a priori. This uncertainty can be treated as uniform randomness. Due to this randomness, it is difficult to determine if a marker on an object is visible to a camera

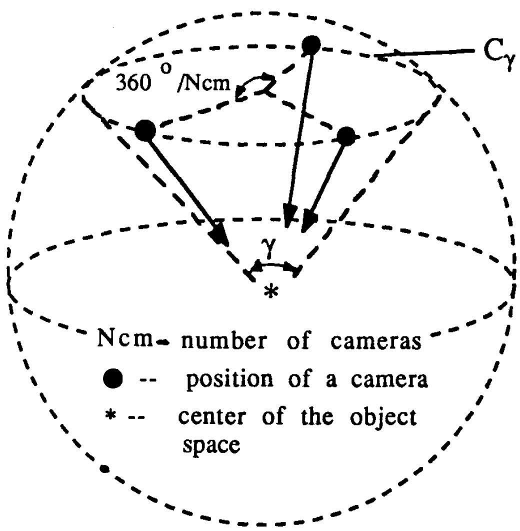

Figure 1: Locations of the Cameras

located at some place on the location-sphere. To facilitate the calculation, this uncertainty is treated in the reverse sense. In other words, the orientation of the viewed object is defined as fixed, and the locations of the cameras are defined to be on one of the facial circles of the location-sphere parallel to the supporting plane of the object. The cameras are randomly yet uniformly distributed on this circle. The only variable which determines the camera locations is then the center-angle of the sphere defined by that circle. Let the center-angle be denoted as γ, and the circle be denoted as C_γ (Figure 1).

Therefore, for an angle γ, the complete set of possible camera locations is defined by the full circle C_γ. A marker on the object may be visible from some of these locations. To evaluate the geometric visibility of the markers, it is necessary to determine the parts of C_γ from which a marker is visible. It should be noted that a 3D semi-infinite volume is associated with each marker, which will be referred to as the ***visible-space*** of a marker. The boundary of a marker's visible-space is determined by the position of the marker on the object and the geometry of the object in the neighborhood of the marker. An example visible-space is shown in Figure 2. Once the position and orientation of an object is fixed, the visible-spaces of all the markers on the object are specified as well. Hence, the intersections of these visible-spaces and the circle C_γ can be determined. These intersections then specify those arcs of C_γ, from which the markers on the object are visible.

3.3 Measures for the detectability of markers

The arcs of C_γ from which the markers are visible can be classified into two different classes: 1) those from which only one marker can be detected by one

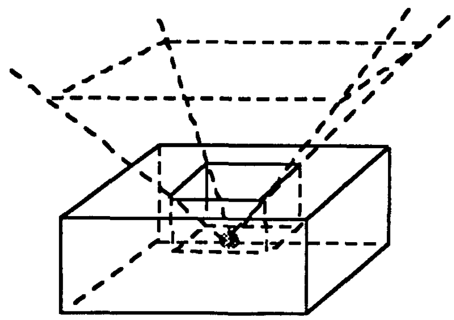

Figure 2: Visible-Space of a Marker

camera, and 2) those from which two or more markers can be detected by either one or more cameras. In the latter case, the goodness of detecting a marker is improved by being able to either have more precise detection, or have the flexibility of selecting one marker from several detected markers. Based on this classification, we define two types of measures:

(1) the probability of detecting one marker, denoted as $G(\gamma, N_{cm})$, defined by the union of those locations on C_γ from which at least one marker is detectable:

$$G(\gamma, N_{cm}) = \sum_{I=1}^{N_{obj}} P_n(I)[\sum_{J=1}^{N_{st}} P(A_J) \times \frac{1}{2\pi}\int_0^{2\pi} \delta_i d\theta] \quad (2)$$

where $P_n(I)$ is the frequency of occurrence of the Ith object for N_{obj} objects, $P(A_J)$ is the occurrence-probability of the Jth stable-orientation for N_{st} stable-orientations. The integrand δ_i in Equation (2) is defined as follows:

$$\delta_i = \begin{cases} 1, & \theta \in VS \\ 0, & \theta \in IVS \end{cases} \quad (3)$$

where VS specifies the set of arcs from which at least one marker is detectable, and IVS on the other hand specifies the set of arcs from which no marker is detectable. These arcs are functions of γ and N_{cm}, and their determination is discussed in the Appendix.

(2) The second measure, denoted as $H(\gamma, N_{cm})$, is defined by the union of those arcs from which two or more markers are detectable, and can be expressed as follows:

$$H(\gamma, N_{cm}) = \sum_{I=1}^{N_{obj}} P_n(I)[\sum_{J=1}^{N_{st}} P(A_J) \times \frac{1}{2\pi}\int_0^{2\pi} \delta_j M d\theta] \quad (4)$$

The integrand δ_j in Equation (4) is defined as follows:

$$\delta_j = \begin{cases} 1, & \theta \in vso \\ 0, & \theta \in vss \end{cases} \quad (5)$$

where vss specifies those members of VS from which only one marker can be detected by one camera, and vso specifies the remaining part of VS from which two or more markers can be detected by one or more cameras. When two or more markers can be detected (by either one or more cameras), each camera-marker pair represents the detection of one relative location of the marker. The integrand M in Equation (4) is equal to the number of camera-marker pairs.

The two measures G and H can be treated in different ways. They could be combined into one criterion, if they are considered as equally important. They could also be treated separately and hierarchically, if they are considered to have different importancy. Our treatment on them is presented in the next section.

3.4 Modification of G and H

In our on-going research on the marker-detection process, we have observed that if the angle, denoted as α, between the camera's optical axis and a marker's normal is not within a suitable range, the calculation of the marker's location can be very sensitive to small changes of that angle. Therefore, from a practical point of view, the definitions of the two measures G and H are modified to take into account the effect of this angle. This effect, denoted as ω, is defined as follows:

$$\omega(\alpha) = \begin{cases} sin\alpha & 0^\circ \le \alpha < 25^\circ \\ sin\alpha + 0.15 & 25^\circ \le \alpha < 30^\circ \\ 1 & 30^\circ \le \alpha < 75^\circ \\ 1 - 0.5 sin\alpha & 75^\circ \le \alpha < 85^\circ \\ 1 - sin\alpha & 85^\circ \le \alpha \le 90^\circ \end{cases} \quad (6)$$

The modified G and H are then accordingly defined as:

$$G' = \sum_{I=1}^{N_{obj}} P_n(I)[\sum_{J=1}^{N_{st}} P(A_J) \times \frac{1}{2\pi}\int_0^{2\pi} \delta_i \omega_m d\theta] \quad (7)$$

and

$$H' = \sum_{I=1}^{N_{obj}} P_n(I)[\sum_{J=1}^{N_{st}} P(A_J) \times \frac{1}{2\pi}\int_0^{2\pi} \delta_j \sum_{k=1}^{M} \omega(\alpha_k) d\theta] \quad (8)$$

where,

$$\omega_m = \begin{cases} \omega(\alpha) & \text{if } \theta \in vss \\ \max\{\omega(\alpha_k), k = 1, M\} & \text{if } \theta \in vso \end{cases} \quad (9)$$

In our proposed solution, the overall detectability of markers (i.e., J_{dt} in Equation (1)) is evaluated by combining these two modified measures:

$$J_{dt}(\gamma) = G' + \lambda H' \quad (10)$$

where λ is a weighting factor. As it can be noticed, the value of λ would affect the results of the optimization. Generally, the selection of λ is not a easy task due to the unpredictablility of the G' and H' curves. However, from the definition of J_{dt}, it is quite straightforward that a very large λ would allow H' to dominate J_{dt}, and a very small λ would allow G' to dominate J_{dt}. If a user wishes to have a system where multiple markers are visible to the camera(s), a high value of λ can be chosen. Otherwise, a low value of λ (or $\lambda = 0$) would favor G' as the dominating part in J_{dt}, which implies that the detection of a single marker is sufficient for the task at hand.

The original function G defined in equation (2) is still employed to evaluate the feasible range of the variable γ for the maximization of J_{dt}, as expressed below:

$$\Gamma \supseteq \{\gamma_i | G(\gamma_i, N_{cm}) \geq p^*\} \quad (11)$$

where p^* is the user chosen threshold on the probability of detecting one marker.

3.5 Optimization algorithm

A multi-mode global-extremum optimization algorithm [6] was used to determine the optimal angle γ^*, for maximum J^*_{dt}, given a number of cameras, N_{cm}. The two-level search process starts from $N_{cm} = 1$. For each N_{cm}, the active range of the variable γ is determined to satisfy the constraint imposed on the minimum acceptable probability of detecting at least one marker. If this feasible range is null, N_{cm} is incremented by 1, and the process of determining the range of γ is repeated. If a non-null feasible range of γ exists, the maximum J_{dt} is sought within this range. The resulting γ^*, which yields the maximum J_{dt}, therefore defines the desired optimal camera location(s), and the corresponding N^*_{cm} is the desired optimum.

IV. SIMULATION RESULTS

The method of optimal camera placement presented in Section 3 was implemented in C on a Sun4/60 workstation. The simulation was carried out for a set of four objects, with given marker arrangements (Figure 3) and several values of p^* (the minimum required value for the probability of detecting one marker). The occurrence-frequency of the objects were considered to

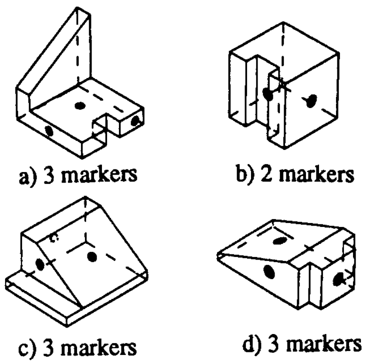

Figure 3: The Objects Used for the Simulation

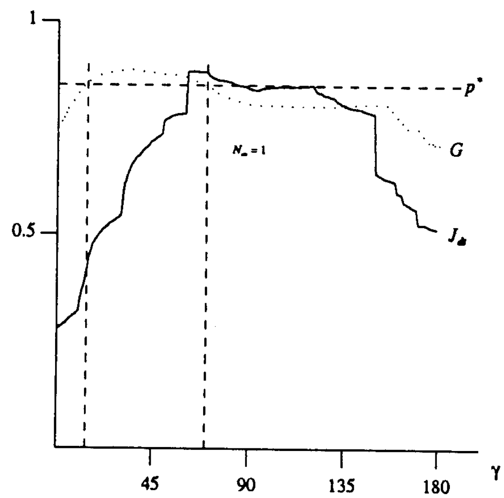

Figure 4: Plots of J_{dt} and G, for $p^* = 0.85$

be as: $P_n(1) = 0.4$, $P_n(2) = 0.2$, $P_n(3) = 0.2$, $P_n(4) = 0.2$. The viewed objects were assumed to rest on a horizontal plane, defining the maximum allowable range of γ to be $0°$ to $180°$. The occurrence-probabilities of the stable-orientations of all the objects were given as well.

The simulation results for maximizing the objective function $J_{dt}(\gamma)$ with $\lambda = 0.1$ are given below for different p^* values:

$p^* = 85\%$,
$G(\gamma, N_{cm} = 1) \geq p^*$, for $\gamma \in \Gamma = [14°, 70°]$, yielded $\gamma^* = 63°$ and $N^*_{cm} = 1$, (Figure 4).

$p^* = 95\%$,
$G(\gamma, N_{cm} = 2) \geq p^*$, for $\gamma \in \Gamma = [16°, 180°]$, yielded $\gamma^* = 63°$ and $N^*_{cm} = 2$, (Figure 5).

$p^* = 100\%$,
$G(\gamma, N_{cm} = 3) = p^*$, for $\gamma \in \Gamma = [27°, 50°]$, yielded $\gamma^* = 50°$ and $N^*_{cm} = 3$, (Figure 6).

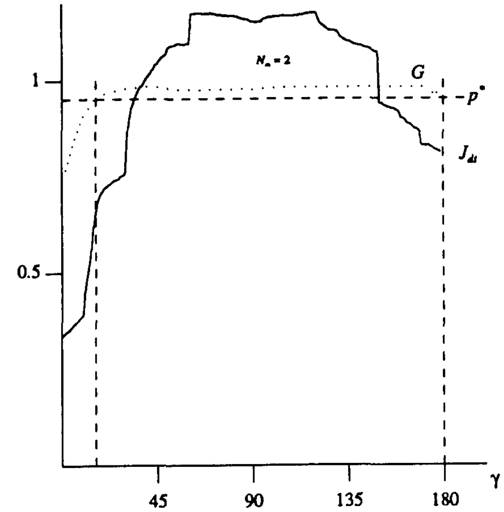

Figure 5: Plots of J_{dt} and G, for $p^* = 0.95$

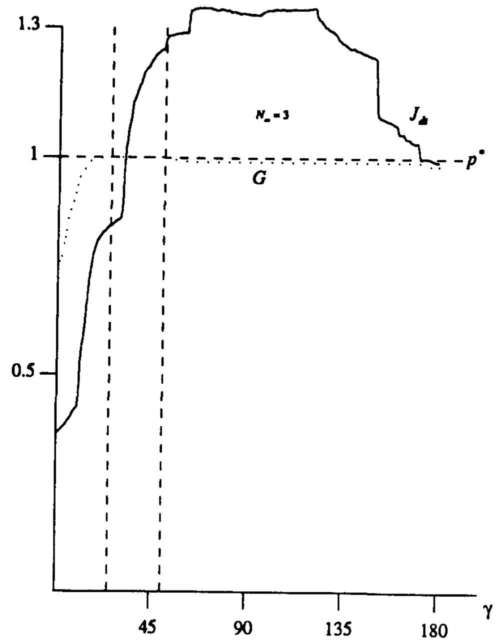

Figure 6: Plots of J_{dt} and G, for $p^* = 1.00$

We have also carried out some tests on the selection of λ. Some differences in the resultant γ^* were observed for different λ values. It was also observed, however, that the differences in the value of J_{dt} corresponding to those different γ^*'s were not significant, for our set of example objects.

V. CONCLUSIONS

The concept of pre-marking the objects is the essence of simplifying the image analysis process. In the context of a vision system which utilizes pre-marking and active sensing for 3D-object-recognition, the detection of a marker in a scene is important to the success of the recognition task. In this paper, we described a method of determining the optimal initial camera locations for a multi-camera marker-detection environment. A two-level optimization is performed with respect to the detectability of the markers, where the minimization of the number of cameras is of primary concern.

References

[1] Safaee-Rad, R., Benhabib, B., Smith, K.C., and Zhou, Z., "Pre-Marking Methods for 3D Object Recognition" *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics*, Vol. II, pp. 592-595, Cambridge, USA, Nov. 1989.

[2] Cameron, A. and Durrant-Whyte, H., "A Bayesian Approach to Optimal Sensor Placement" *The International Journal of Robotics Research*, Vol. 9, No. 5, pp. 70-88, 1990.

[3] Hager, G.D., **Task-Directed Sensor Fusion and Planning**, pp. 137-152, Kluwer Academic Publishers, USA, 1990.

[4] Tsai, R.Y. and Tarabanis, K., "Model-Based Planning of Sensor Placement and Optimal Settings" *SPIE Vol. 1198, Sensor Fusion II: Human and Machine Strategies*, pp. 75-89, 1989.

[5] Yi, S., Haralick, R.M. and Shapiro, L.G., "Automatic Sensor and Light Source Positioning for Machine Vision" *IEEE 10th Int. Conf. on Pattern Recognition*, pp. 55-59, Atlantic City, USA, 1990.

[6] Babii, A.N., "Algorithm for Finding the Value of the Global Extremum of a Function of Several Variables with a Given Accuracy" *Cybernetics*, Vol.14, No. 5, pp. 699-703, 1978.

APPENDIX

Determination of VS and IVS

Given the objects' geometric models and the positions of the markers on each object, the boundary of each marker's visible-space can be expressed by a group of algebraic equations. For a certain γ, the intersection of C_γ and the visible-space of a marker can be determined by solving a system of equations which represent the circle C_γ and the visible space through analytic-geometry. The details of the calculations are omitted here.

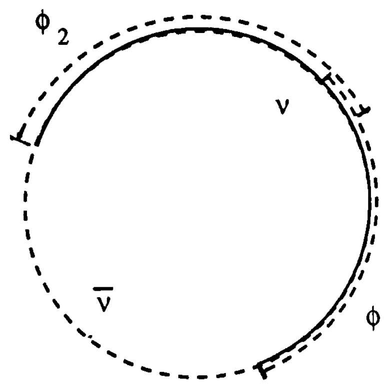

Figure 7: Spatial Visible/Invisible Arcs of C_γ

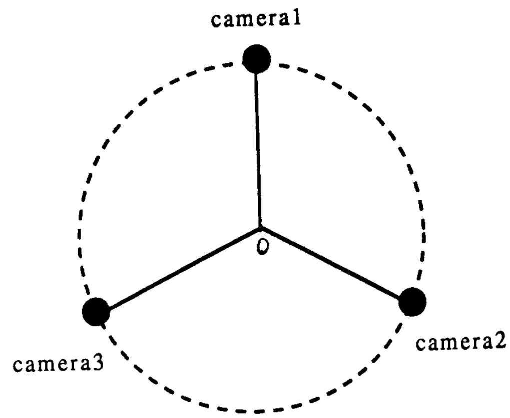

Figure 8: The Camera Pattern for Determining VS/IVS from the Spatial Visible/Invisible Arcs

Once the intersections of C_γ and the visible-spaces are determined, the circle C_γ is decomposed into visible arcs (inside the visible-space) and invisible arcs (outside the visible space), with respect to each marker. Let ϕ_i denote the set of the arcs of C_γ, where marker i is visible, and η_i denote the rest of C_γ, where marker i is not visible. By determining the union of the visible arcs for all the markers, C_γ is decomposed into two sets of arcs with respect to one stable-orientation: the set of visible arcs, denoted as ν, and the set of invisible arcs, denoted as $\overline{\nu}$.

$$\begin{aligned} \nu &= \cup\, \phi_i, \quad i = 1, k \\ \overline{\nu} &= \cap\, \eta_i, \quad i = 1, k \\ C_\gamma &= \nu + \overline{\nu} \end{aligned}$$

These two sets of arcs represent the spatial visible/invisible positions with respect to one stable-orientation. An example of the spatial visible/invisible arcs is shown in Figure 7. Then, for a certain number of cameras, the circle C_γ is further decomposed into the actual detectable and undetectable arc sets. The uniform distribution of the cameras is applied to determine these two sets. Each member of the set of detectable arcs is associated with the information that states "which of the markers (specified by marker's identity number) can be observed by which of the cameras (specified by camera's identity number)". A member of the set of undetectable arcs specifies the position from which none of the cameras can observe any of the markers. Let VS denote the set of detectable arcs and IVS denote the set of undetectable arcs, then:

$$C_\gamma = VS + IVS$$

The determination of VS and IVS can be performed by the following approach: Define a planar pattern on C_γ by connecting the position of each of the uniformly distributed cameras to the center of C_γ. For instance, such a pattern for the case of three cameras is shown in Figure 8. Let this pattern rotate from 0 to 2π. Then, for each moment of the rotation, determine which of the cameras coincides with the set of spatial visible arcs ν. All coinciding points belong to the set VS, while all others belong to IVS. For one camera, $VS = \nu$ and $IVS = \overline{\nu}$.

Among those points on C_γ which belong to VS, some are referred to as "members of vss" if from their positions only one marker can be detected by one camera, and the others are referred to as "members of vso" from which two or more markers can be detected by one or more cameras. That is:

$$VS = vss + vso\ .$$