Optimal Camera and Marker Arrangements for an Active-Vision System

X. He B. Benhabib K. C. Smith R. Safaee–Rad

Computer Integrated Manufacturing Laboratory
Department of Mechanical Engineering
University of Toronto
5 King's College Road
Toronto, Ontario
Canada M5S 1A4

Abstract **A new active-vision system for 3D object recognition in robotic assembly workcells is being developed in our Laboratory. In this system, image analysis is simplified by pre-marking the objects with artificial *markers* and utilizing a mobile camera to acquire views of the objects pre-specified by markers. Hence, the objects are represented by a small set of 2D perspective projections. In this context, an important problem to be addressed is the determination of the number of markers and their locations on the objects, and of the placement of the mobile camera in order to guarantee the detection of at least one marker. In this paper, we present a solution to the optimal arrangement of markers and camera location(s), given the objects' solid models and a set of constraints.**

I. INTRODUCTION

In order to develop an efficient 3D object-recognition method for vision-based robotic assembly, a new active-vision system is under development in the Computer Integrated Manufacturing Laboratory at the University of Toronto [1]. In this system, the 3D objects are represented by a small number of their 2D distinct perspective projections. These views are referred to as *standard-views*. Each view is specified by a local surface normal, referred to as *standard-view-axis*. For successful recognition, the input image of an object must be acquired from one of its standard-view-axes. Hence, a mobile camera is employed so that a standard-view can be acquired by aligning the camera's optical axis with one of the standard-view-axes. The matching process is thus performed between the acquired 2D standard-view image of an unknown object and the library of 2D standard-views of a set of objects.

As mentioned above, to enable the acquisition of a standard-view, the standard-view-axes must be predefined as well. This can be accomplished by pre-marking the objects, that is, adding nonfunctional features onto the objects. The nonfunctional features that we are proposing, referred to as *markers*, are flat circles. Hence, the normal of a marker passing through the center of the marker circle defines a standard-view-axis.

Pre-marking schemes for the simplification of image analysis have been proposed and used in different contexts, such as: with metal-marking dyes to enhance image contrast for industrial visual inspection of complex assemblies [2]; and, as a standard pattern for the accurate relative position determination and verification by a mobile camera [3]. Although the idea of using markers for object recognition was first proposed more than ten years ago, the study and application of this technique for industrial use is still open for investigation.

In order to utilize the proposed recognition strategy, a set of standard-views for the objects must be selected and stored in a standard-view database, and, accordingly the objects must be pre-marked in order to define the local references for the acquisition of the standard-views. The selection of a specific set of standard-views, in fact, is equivalent to the pre-marking of objects (i.e., the determination of the number of markers and their locations on the objects), since a standard-view is the projection of an object from a viewpoint indicated by the position of a particular marker. There are two aspects of the determination of the number and locations of the markers: one is to facilitate the process of marking the objects and the matching process, the other is to facilitate the detection of markers.

One of the requirements for successful recognition is the existence of detectable markers in the scene. Since the success of detecting a marker's existence, as a part of the marker-detection process, depends on the location of the mobile camera as well, the placement of the camera's initial locations is to be determined along with the arrangement of the markers on the objects. By "initial locations" it is implied the maximum required suc-

0-7803-0720-8/92 $3.00 ©1992 IEEE

cessive locations of the mobile camera for the detection of at least one marker. Therefore, two of the problems to be solved in the context of the proposed vision system are the determination of an optimal camera placement and an optimal pre-marking of a set of objects, under a set of constraints.

The optimal camera placement and the optimal object pre-marking problems were previously handled individually [4,5]. However, as mentioned earlier, these two issues can be seen as interdependent and to be solved together. In this paper, aspects of our previous work are combined and extended. The objective of the problem addressed here is to represent the 3D objects by a small and optimal set of 2D views, and at the same time, to determine the optimal initial locations of a mobile camera for satisfying a desired degree of certainty in detecting the markers.

II. ASSUMPTIONS

The following assumptions are made: (1) The camera is located on the surface of a virtual sphere centered at the origin of a predefined frame referred to as a *scene-frame*, while the camera's optical axis points to the center of the sphere where the viewed object is normally located; (2) The field-of-view of the camera is large enough to include the viewed object; and, (3) There exist only a finite number of resting-positions for each object, with each object having a known arrival-frequency into the scene, and each resting-position having a known occurrence-probability.

The sensitivity analysis, discussed in Section VI, will show that the optimization results are not very sensitive to these assumed parameters/relations.

III. DEFINITIONS

A. Representation of the camera's initial location(s)

In this paper, the *initial* locations of the camera are defined as the successive locations for the mobile camera from which to search for the first detectable marker. As mentioned earlier, the camera's location is assumed to be located on the surface of a virtual sphere, referred to as the camera's *location-sphere*.

For an object observed in the scene, its rotational orientation about the normal of the support-plane is completely random. This randomness makes it difficult to determine whether a marker on the object is visible to the camera located somewhere on the location-sphere. To facilitate the calculation, this randomness is treated here in the reverse way, where the orientation of the viewed object is considered to be fixed, and the mobile camera is defined to be randomly located on one of the circles on location-sphere surface whose plane is parallel to the support-plane of the object, Figure 1. Correspondingly, the center-angle of the sphere, γ, defined by the circle C_γ, is the variable that specifies the camera's orientation. For any γ angle, the mobile camera will be randomly placed on C_γ. One approach to guaranteeing the detection of the markers, for an angle γ, is to allow the camera to have more than one position on the circle. Then, if the camera when placed randomly on a circle does not detect any marker, it can be moved to another position. For convenience, for any γ angle, a number of uniformly-spaced positions for the camera are defined on the circle C_γ. The total number of such positions is referred to as the *number of initial camera positions*. Placing the camera at at most this number of positions on a circle, the desired degree of certainty of marker detection is guaranteed.

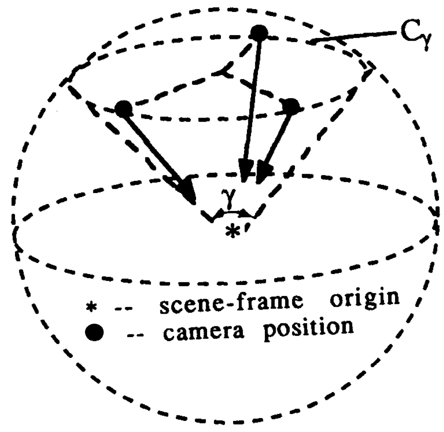

Figure 1. Camera's Location-Sphere

Thus, in summary, the initial locations of the camera are represented by an angle γ (which specifies the orientation of the camera) and the corresponding number of initial camera positions, which are randomly, yet uniformly, spaced).

B. Detectability of markers

The detectability of markers, in general, includes two issues: both geometric and optical detectability. In this paper, detectability refers to the geometric detectability of the markers. Optical detectability, which is affected by factors such as illumination, image resolution, etc., is not considered.

Geometric detectability of markers relies on some measure of the possibility of detecting a marker. In [4], we defined two types of measures of detection in terms of geometric visibility of the markers: 1) the chance of detecting at least one marker (denoted by G'), and 2) the chance of detecting two or more markers (denoted by H'). The introduction of these two types of measures is motivated by the requirement of detecting at least one marker, and in some cases more than one marker, for successful identification of an object in the proposed vision system.

The overall detectability of markers (denoted as J_{dt}) is hence defined as the combination of the two measures:

$$J_{dt} = G' + \beta H' = \sum_{i=1}^{N_{obj}} p_n(i)[\sum_{j=1}^{N_{st}(i)} (p_j [g' + \beta h'])] \quad (1)$$

where β is a weighting factor, N_{obj} is the number of the objects; $N_{st}(i)$ is the number of resting-positions for the object i; $p_n(i)$ is the known arrival-frequency of object i; and p_j is the given occurrence-probability of resting-position j.

With meanings that are similar to G' and H' respectively, functions g' and h' are calculated with respect to each resting-position. g' is calculated on a probabilistic basis by finding the arcs of C_γ from which at least one marker is visible to the camera; and h' is calculated by finding those arcs from which two or more markers are visible. Mathematical details of the definition and calculation of g' and h' are presented in [4].

C. Distinctiveness of a set of views

The standard-views in the database are intended to be as dissimilar from one another as possible. The goodness of a set of views is hence referred to as its *distinctiveness*, denoted by F. For a set of views, distinctiveness is dependent on the number of the views and the shape complexities of these views.

Generally, two views can be distinguished by the distance between what characterizes their shapes (such as feature-vectors or other descriptors). In this paper, the distance is evaluated by an average-pointwise-distance measure of their Angle-Of-Sight (AOS) signatures. The AOS signature is a boundary-based shape descriptor [6].

The properties of a set of views in which we are interested are: the average distance amongst all the views, and the minimum distance between any two views. Therefore, F is defined by the weighted sum of the average distance and the minimum distance between all possible pairs of views in the set. It is formulated as follows:

$$F = \alpha_1 (\frac{1}{n}\sum_{i=1}^{n} d_i) + \alpha_2 (min \{ d_i, i=1,n \}) \quad (2)$$

where d_i is the distance between the ith pair of views, n is the total number of possible pairs for a set of views, and α_1 and α_2 are weighting factors.

IV. PROBLEM STATEMENT

Motivated by the requirement of efficiency and the goal of successful recognition, the objectives for an optimal camera placement and object pre-marking are:

- the number of initial camera positions is as small as possible, so as to reduce number of camera moves in the search for the first detectable marker;
- the number of markers is as small as possible, so as to reduce the size of the standard-view database;
- the distinctiveness of the set of standard-views is as high as possible; and
- the detectability of the markers is as high as possible.

Amongst the four criteria above, the most important one is the number of camera positions, since the time spent on moving the mobile camera could easily be longer than the time spent on comparing two views. The second criterion of importance is the number of markers, since fewer markers requires that fewer surfaces of objects be premarked and fewer view-comparisons be made in the matching process. As for the detectability of markers and the distinctiveness of the standard-views, they can be either combined into one criterion (which is the approach chosen in this paper), or optimized hierarchically if one is preferred over another. The four variables in this problem are as follows: the number of initial camera positions (denoted by N_{cm}); the number of markers (denoted by N_{mk}); the set of 2D views (denoted by S) whose number of elements is equal to N_{mk} with the shape of each element determined by the location of each marker; and, the orientation of the camera (γ).

The constraints for this problem are: the probability of detecting at least one marker is 100%; the distance between any two standard-views is greater than a given threshold; the number of feasible standard-views for each resting position is at least two; and, the number of initial camera positions and the number of markers do not exceed given upper bounds.

A. Three-Level Optimization

The above hierarchical multi-objective problem can be defined as a three-level optimization formulated as:

$$\begin{array}{ll} \text{Minimize} \quad N_{cm} & \\ \quad \text{Minimize} \quad N_{mk} & \\ \quad\quad \text{Maximize} & Q(N_{cm}, N_{mk}, S, \gamma) = \lambda_1 J_{dt} + \lambda_2 F \\ \text{Subject to} & \gamma \in \Gamma = \{\gamma \mid G(N_{cm}, N_{mk}, S, \gamma) = 1\} \quad (3) \\ & d_i > D_0, \; i = 1, n \\ & S \cap C_{ij} \geq 2, \; j = 1, N_{st}(i); \; i = 1, N_{obj} \\ & N_{mk} < K \\ & N_{cm} < M \end{array}$$

where Q is the objective function for the innermost level of the optimization, λ_1 and λ_2 are weighting factors, Γ represents the feasible region of the variable γ, G is the probability of detecting at least one marker, [4], D_0 is a given threshold for the distance between any two standard-views, C_{ij} is the set of visible views for the jth resting-position of object i, and K and M are given upper bounds for the number of markers and camera positions respectively.

Correspondingly, given a set of objects and a mobile camera, the problem of optimal object pre-marking and optimal camera placement can be solved with three nested-loops in which the outermost loop updates the number of camera positions iteratively, the intermediate loop updates the number of markers iteratively, and the innermost loop searches for the maximum of its own criterion so as to yield the best S and γ.

It should be pointed out that this optimization problem is combinatoric in nature. There exists a large number of candidates/combinations for the variable S, and a correspondingly excessive demand for computation time. The most difficult part of solving this optimization problem is how to explore a limited range of S and yet preserve the globality of one optimal solution.

Generally, any flat surface of an object is a candidate for determining an element of S. To eliminate non-feasible flat-surface candidates, a set of heuristic rules are applied such that: 1) a candidate surface is not self-obscured; 2) the surface is sufficiently large to hold a marker; and 3) if there are several parallel surfaces for an object, only the largest one is chosen. Unfortunately, too many candidates of S usually remain after the application of these heuristics.

However, by studying the specific constraints of this optimization problem, an algorithm which significantly reduces the number of feasible combinations has been derived. The algorithm is based on the following two properties:

(1) Let S_i (i=1, N_{obj}) consists of all the members of S that belong to the ith object. Then, the constraint $S \cap C_{ij} \geq 2$ is satisfied if and only if $S_i \cap C_{ij} \geq 2$ (j=1,$N_{st}(i)$) are satisfied for each individual object. The basis for such a relationship is that the resting-positions of one object are independent of those of the other objects.

(2) The necessary condition for satisfying the constraint G=1 (with respect to all the objects) is that the probability of detecting at least one marker is 100% for the markers on any individual object.

Therefore, after the heuristic rules have been applied, the final number of feasible combinations can be determined by a process of two-step scanning. Firstly, the constraints $S_i \cap C_{ij} \geq 2$, $d > D_0$, and G=1 are examined for all possible combinations of views for each individual object, yielding a group of candidate combinations for each object. Secondly, the groups of candidate from all the objects are considered together to generate all possible combinations. Then the constraints $d > D_0$ and G=1 are examined for these combinations, and the final set of feasible candidates (of S) is determined.

V. SIMULATION RESULTS

Computer simulations were carried out utilizing the set of objects shown in Figure 2. Their resting-positions considered are shown in Figure 3. A CAD system (namely the I-DEAS package by SDRC) was used to create their solid models and to generate 2D perspective *shaded images* of the objects. Interface programs were developed to retrieve the solid models and 2D images from the CAD system, and to input appropriate information to the optimization program.

By choosing a normalized threshold of $D_0 = 3\%$, upper bounds as $M = 3$ and $K = 24$, and selecting the weighting factors as $\alpha_1 = \alpha_2 = 0.5$, $\beta = 1$, and $\lambda_1 = \lambda_2 = 0.5$, the following results were obtained: $N_{cm} = 1$, $N_{mk} = 13$, $\gamma = 32°$, with the best locations of the markers on the set of objects being as shown in Figure 4.

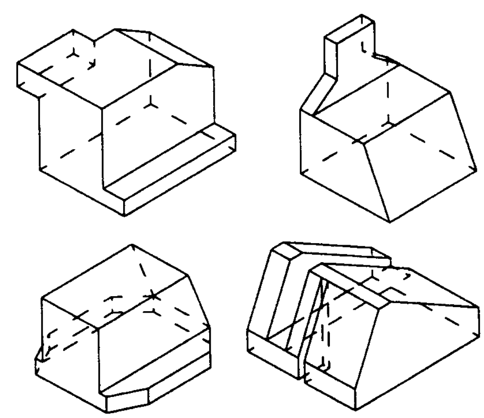

Figure 2. A Set of Objects

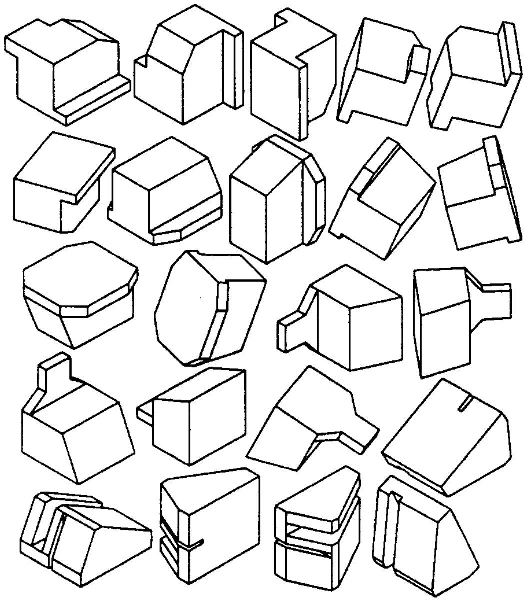

Figure 3. Resting-positions Considered

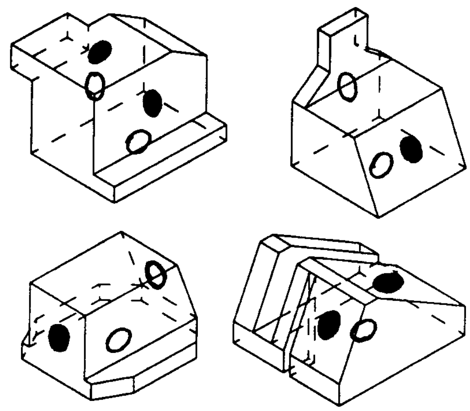

Figure 4. Optimal Marking Results

VI. DISCUSSION -- SENSITIVITY ANALYSIS

In solving the above problems, a few assumptions were made with respect to certain parameters, such as the occurrence-probabilities of the resting-positions of the objects, the center of the camera's location-sphere. A question one may ask is: how would the result of the optimization be affected if the actual parameters were not actually as assumed? Accordingly, we have conducted a sensitivity analysis by studying the effect of inaccuracies of such parameters on the objective function Q defined in Section IV.

The parameters examined were: the occurrence probability of each resting-position (p_{ij}), the center of camera's location-sphere in the scene-frame (denoted as c), position of the object in the scene-frame (denoted as b), the radius of the camera's location-sphere (denoted as R), and the position of the marker's center on a surface (denoted as m).

Considering all five parameter types, the objective function is expressed as:

$$Q = Q(p_{ij} ; c ; b ; R ; m; \boldsymbol{u}) \qquad (4)$$

where $\boldsymbol{u} \supseteq (N_{cm}, N_{mk}, S, \gamma)$.

Generally speaking, sensitivity of a function to its parameters is usually studied through the partial derivatives of the function with respect to those parameters. However, in our case, for most of the parameters, explicit expressions for such partial derivatives are not available. What we substitute is an experimental sensitivity analysis.

The effect of a parameter-estimation inaccuracy on the objective function is defined as:

$$\delta Q = \frac{Q^o - Q'}{Q_M - Q_m} \qquad (5)$$

where Q^o is the desired (optimal) maximum of the function Q determined using the true value of the parameter, Q' is the actual function value determined using the actual parameter value, and Q_M and Q_m are the highest and lowest possible values that Q could reach respectively (used for the purpose of normalization).

For inaccuracies in the five types of parameters, there are two cases to be considered: In the first, the parameters used for determining the optimal solution are inaccurate, while the actual run-time parameters are true values (the inaccuracy in the occurrence-probability p_{ij} is of this type). In the other case, the parameters used for determining the optimal solution are the (desired) true values, while the actual ones may not be accurate (the inaccuracies in c, b, R, and m are of this type).

Therefore, for p_{ij}, the effect of its inaccuracy on Q is calculated using Equation (5), where:

$$Q^o = Q(p^o_{ij} ; \boldsymbol{u}^o) \ , \quad Q' = Q(p^o_{ij} ; \boldsymbol{u}') ; \qquad (6)$$

in which, p^o_{ij} are the true values of the probabilities, $\boldsymbol{u}^o$ is the optimal solution obtained from the global-optimization performed with the true probabilities, and $\boldsymbol{u}'$ is the optimal solution obtained with the given (inaccurately estimated) probabilities.

For c, b, R, and m, on the other hand,

$$Q^o = Q(x^o ; \boldsymbol{u}^o) \ , \quad Q' = Q(x' ; \boldsymbol{u}^o) ; \qquad (7)$$

where x presents any one of the parameters c, b, R, and m; x^o is the desired value of the parameter used in obtaining the optimal solution $\mathbf{u}^o$; and x' is the parameter's actual run-time value.

When there are some uncertainties in all five types of parameters, their total effect on the objective function can be estimated as:

$$\delta Q_T = [\delta Q_1^{\,2} + \delta Q_2^{\,2} + \delta Q_3^{\,2} + \delta Q_4^{\,2} + \delta Q_5^{\,2}]^{1/2} \quad (8)$$

where the subscripts 1 to 5 refer to the parameters p_{ij}, c, b, R, and m, respectively, [7].

Computer simulations were carried out by introducing a range of levels of inaccuracy in these parameters. Some of the sample results are shown in Table 1, where the weighting factor λ_2 is not listed, but assigned as $\lambda_2 = 1 - \lambda_1$; δp_{ij} is the average of the inaccuracy of every p_{ij} (which is defined as the difference between the nominal and actual values of a p_{ij} divided by the nominal value); δc (or δb) is defined as the distance between current and assumed positions of the location-sphere's center (or of the object) divided by half of the maximum dimension of the objects; δR is the difference between the assumed and actual radii of the location-sphere divided by the assumed radius; δm is the average of each marker's *shifting-factor* which is defined as the distance between the assumed and actual centers of a marker divided by the radius of an envelope-circle of the surface on which the marker is placed. The results show that the optimization is not very sensitive to the uncertainties in these parameters (e.g., if the uncertainties in the parameters are under 12%, their total effect is within 5%).

Table 1. Sample Results of the Sensitivity Analysis

Parameters	λ_1	δQ_1	δQ_2	δQ_3	δQ_4	δQ_5	δQ_T
δp_{ij}=11%, δc=9%	0.5	0%	2%	2%	1%	1%	3%
δb=9%, δR=10%, δm=11%	0.9	0%	3%	3%	2%	2%	5%

VII. CONCLUSIONS

The 3D object-recognition system developed in our Laboratory utilizes a mobile camera to gain system flexibility, and a pre-marking scheme to simplify the image-analysis process so as to gain system efficiency. In this paper, we have presented a method for the solution of a fundamental problem embodied in the proposed vision system, namely, the determination of the optimal arrangement of markers on a given set of objects, and the optimal placement of a mobile camera to satisfy a desired degree of certainty in detecting the markers. The methods developed for solving the optimization problem provide a mathematical basis for the use of a specific pre-marking scheme in 3D object-recognition, and for the solution of the sensor-placement problem in terms of geometric detectability of special features, in a quite general way.

REFERENCES

[1] Safaee-Rad, R., "An Active-Vision System for 3D-Object Recognition in Robotic Assembly Workcells", *Ph.D. dissertation*, University of Toronto, Toronto, Canada, Sept. 1991.

[2] Bodziak, T.A., "Using UV Metal-Marking Dyes in Industrial Vision Inspection of Complex Assemblies", *Proc. of the Vision'86*, pp. 9-31 to 9-39, Detriot, USA, 1986.

[3] Kabuka, M.R. and Arenas, A.E., "Position Verification of a Mobile Robot Using Standard Pattern", *IEEE J. of Robotics and Automation*, Vol. RA-3, No. 6, pp. 505-516, 1987.

[4] He, X., Benhabib, B., Smith, K.C., and Safaee-Rad, R., "Optimal Camera Placement for an Active-Vision System", *Proc. of IEEE Int. Conf. on Systems, Man, and Cybernetics*, Vol. I, pp. 69-74, Charlottesville, Oct. 1991.

[5] He, X., Benhabib, B., and Smith, K.C., "Optimal Object Modeling for an Active-Vision System", *Proc. of Canadian Conference on Electrical and Computer Engineering*, Vol. I, pp. 4.3.1-4.3.4, Quebec, Sept. 1991.

[6] Tchoukanov, I., Safaee-Rad, R., Benhabib, B., and Smith, K.C., "A New Boundary-Based Signature for Industrial-2D-Shape Analysis", *IEEE/RSJ Int. Conference on Intelligent Robots and Systems (IROS '92)*, pp. 1030-1037, Raleigh, North Carolina, July 1992.

[7] Moffat, R. J., "Contributions to the Theory of Single-Sample Uncertainty Analysis", *J. of Fluids Engineering*, Vol. 104, pp. 250-260, June 1982.