CAD-Based Off-Line Planning for Active Vision

Xiaoyu He,* Beno Benhabib,† and Kenneth C. Smith
Computer Integrated Manufacturing Laboratory (CIMLab)
Department of Mechanical Engineering
University of Toronto
5 King's College Road
Toronto, Ont., Canada M5S 1A4

Received April 14, 1994; revised January 9, 1995; accepted May 9, 1995

A novel active-vision system for 3-D object recognition was developed. The vision system combines object pre-marking and active sensing. Therein, an object is modeled by a few of its 2-D perspective (standard) views, each with a corresponding viewing axis defined by a marker placed on the object. The operation of the active-vision system is characterized by its off-line planning and on-line recognition phases. The primary objective of this article is to address two issues in regard to the former: optimal object pre-marking and optimal camera placement. The optimal object pre-marking problem is defined as the determination of the minimum number of markers, and their best locations on a given set of objects to yield maximum "distinctiveness" for the 2-D standard-views defined by these markers. The optimal camera-placement problem targets minimization of camera movements in detecting the markers. Following a brief description of the proposed on-line recognition process, the techniques developed for the off-line planning stage are described in this article. For experimental verification purposes, a prototype of the proposed active-vision system was also implemented and is described herein. © *1995 John Wiley & Sons, Inc.*

†To whom all correspondence should be addressed.
*Currently with Pratt & Whitney, Canada.

Journal of Robotic Systems 12(10), 677–691 (1995)

CCC 0741-2223/95/0100677-15

3D 物体認識用の高性能アクティブビジョン・システムを開発した。このビジョン・システムでは、物体のプリマーキングとアクティブ・センシングを組み合わせている。そして、幾つかの 2D 透視（標準）視野によって、物体をモデル化する。各 2D 透視視野に対応する視野軸は、物体上に設定されるマーカーによって決まる。
アクティブビジョン・システムの動作特性は、そのオフライン計画とオンライン認識の段階によって決まる。この発表の主な目的は、前述の事に関連する 2 つの問題、最適物体プリマーキングと最適カメラ配置について説明することである。
最適物体プリマーキング問題は、最小数のマーカーの決定と、これらのマーカーによって決まる 2D 標準視野で最大の「識別能力」を発生させるための物体集合上の最適マーカー配置決定として定義される。最適カメラ配置問題では、最少のカメラ動作でマーカーを検出することが目的となる。
提案したオンライン認識処理の簡単な説明に続いて、オフライン計画のために開発された技法を説明する。実験確認の目的で、提案したアクティブビジョン・システムのプロトタイプがすでに動作しているので、それについて説明する。

1. INTRODUCTION

The intelligence of robotic workcells can be increased by the employment of machine vision. The essential functions that a machine-vision system should provide are: the identification of workpieces/subassemblies, and the estimation of their 3-D locations. This information would be subsequently used by a robotic arm to move its end-effector to a pre-grasping location, where proximity sensors take over the control of the robot motion toward the final approach to a grasping location.

Most of the past research in machine vision and computer vision has involved the analysis of images acquired with a passive camera. With active perception, however, the location of the camera can be controlled.[1,2] Therefore, if certain *standard* viewing axes were pre-defined and easily detected, the camera could be manipulated for the acquisition of *standard* 2-D images. In this way, the matching process would be performed amongst 2-D views.

Based on the above premise, an active-vision system based on a new model-based object-recognition method was developed in our laboratory.[3] The proposed system combines *active visual sensing* and *pre-marking* of workpieces. In the context of this project, a set of major problems were addressed: object pre-conditioning via marking for specifying the set of 2-D standard-views necessary to represent the objects; placement of the camera in terms of satisfying the requirements of marker detection; feature extraction and matching; and 3-D location-estimation of the object. Amongst these issues, the object pre-marking and the camera placement problems are addressed in this article. The former is to determine a minimum number of markers and their best locations on the objects, so that the *standard* 2-D views corresponding to these markers will model the objects optimally. For the latter, the best location(s) of the camera is to be determined, so that a marker (or more markers) on an object oriented arbitrarily in the scene will always be detected.

2. THE PROPOSED ACTIVE-VISION SYSTEM

2.1. Overview

The modeling or representation of 3-D objects can be divided into two categories: (1) object-centered representations,[4] and (2) multiview representations.[5,6] For the active-vision system developed in the CIMLab, the modeling method belongs to the multiview-representation category. However, an object is modeled via pre-marking by only a small number of its 2-D perspective views, referred to as its *standard-views*. Each standard-view is associated with a corresponding *standard-view-axis*. For successful recognition purposes, the input image of an object must be one of its standard-views. Thus, for the acquisition of a standard-view, the optical axis of a mobile camera has to be aligned with one of the standard-view-axes. The matching process is thereafter performed between the acquired 2-D standard-view of the object and the library of 2-D standard-views of a set of objects.

Previous use of markers in machine vision include: enhancing image contrast for the inspection of complex assemblies,[7] relative-position determination for mobile robots,[8] object identification,[9] and location estimation.[10] In the context of our scheme, object pre-marking serves three purposes: (1) to specify a set of 2-D views by which to represent the objects (i.e., standard-views); (2) to define a local surface normal (i.e., standard-view-axis), which can be determined from the shape of a marker's image; and, (3) to convey local 3-D orientation and position information concerning a surface of a viewed object (such information can be subsequently used for 3-D location-estimation of the object).

Active vision, on the other hand, serves only the purpose of facilitating the acquisition of standard-views of the objects. In the case where the first standard-view image is insufficient to allow the object to be recognized, additional (standard-view) images would have to be acquired. In this context, optimal camera placement is a paramount issue.

The optimal placement problem of cameras/sensors has been studied in different contexts: determining the optimal (visual-) sensor locations for performing recognition and localization operations, based on the acquired partial knowledge of the object[11]; determining the optimal placement and optical-settings of vision sensors for satisfying particular image constraints[12]; determining optimal sensor and light-source positions for edge detection[13]; determining successive viewpoints of a camera to acquire *distinguishing features* of objects[14]; and determining the best "next view" in the context of obtaining a complete model of a scene.[15]

2.2. Proposed Vision System

The general operation of the proposed vision system is divided into two principal phases: off-line planning and on-line recognition. To execute the recognition process, the following tasks must be accomplished off-line: (1) determining the optimal number of markers together with the surfaces of the objects on which these markers are to be placed; (2) providing a library of standard-views, preferably via CAD data; (3) determining the optimal *initial location* of the camera and providing a search strategy so that a marker on a viewed object will always be detected; and (4) determining the optimal *secondary locations* of the camera and providing a search strategy so that additional markers can always be detected. These off-line planning issues form the basis of the present article, for which formulations and solution methods will be provided. In the next section, the modules comprising the on-line recognition process are described briefly.

2.3. On-line Recognition

2.3.1. Marker Detection

In the proposed vision system, the markers employed are planar circles of known size. The projection of a marker on the image plane of the camera is thus an ellipse. The five basic parameters of a marker's elliptical image can be estimated by fitting an elliptical curve to a given set of N image-boundary points through the use of a Minimum-Squares-Error (MSE) criterion.[17,18] The elliptical parameters of the image of a marker combined with the given effective focal length of the camera can be used to determine the 3-D location of a marker with respect to the camera frame.

Determining the 3-D orientation of a marker is equivalent to solving the following problem: Given a 3-D conic surface whose vertex point is the center of the camera lens and whose base is the projection of a marker on the image plane (namely, an ellipse), determine the orientation of a plane with respect to the camera frame (F_c), such that the intersection of the plane and the conic surface is a circle. An analytical solution to this problem was obtained[16] by applying a series of frame transformations.

2.3.2. Standard-View Interpretation

The location of a marker, determined as explained here, guides the camera for the acquisition of a corresponding standard-view of the object. Once acquired, this image must be interpreted for object identification. In the context of the proposed active-vision system, the 2-D shape-recognition technique must be *position-*, *rotation-*, and *size-invariant*. A new boundary-based 2-D shape-recognition technique, which has this property, was developed and reported earlier.[20] This technique characterizes a 2-D shape with a 1-D signal derived from the shape by using the *angle-of-sight* (AOS) encoding scheme. Two fixed points are predetermined as follows: Point C is at the center of the 2-D image's characteristic ellipse (CE)[a]; and Point D is on the line that

[a]As shown in Tchoukanov et al.,[20] for any planar shape a special ellipse, referred to as the *characteristic ellipse* (CE), can always be generated by fitting the shape's boundary points into an elliptical curve.

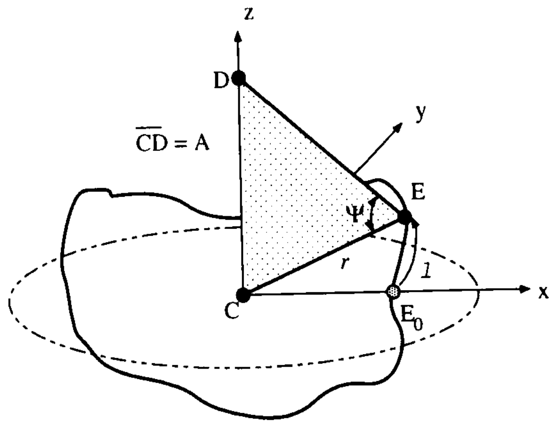

Figure 1. Angle-of-sight encoding scheme.

passes through C and is perpendicular to the plane of the 2-D image (Fig. 1). The distance between points C and D is equal to the length of the major axis of the CE. Therefore, the AOS signature of a 2-D shape is defined as a 1-D signal, $AOS = \Psi(l)$, where l is the arc length between the moving point E and a starting point E_0.

2.3.3. 3-D Location-Estimation of a Viewed Object

The problem of determining the 3-D location of an object is equivalent to determining the 3-D location of a pre-defined body frame of the object, F_0, with respect to a reference frame in the scene, F_u. The solution method[19] is based on the definition of a local frame for each marker, F_m. At run-time, the relation between F_c and an F_m can be measured. Once the object is identified, the relation between its F_0 and the F_m can be retrieved from a precompiled library of transformation matrices, and the location of the object with respect to F_u can be determined.

3. OPTIMAL PRE-MARKING OF OBJECTS

3.1. Problem Formulation

The pre-marking of objects involves the simultaneous determination of both the number of markers and their locations on *all* the objects. Correspondingly, one objective would be to directly minimize the number of markers and therefore minimize the size of the standard-view database. Another objective would be to maximize the distinctiveness of the standard-views defined by the selection of marker locations.

3.1.1. Distinctiveness of a Set of 2D Views

For a selected set of views, an overall distinctiveness measure is defined herein as the combination of (1) the average distance measure between all possible *pairs* of views, and (2) the distance measure for the specific pair with *most similar views*:

$$F(S, N_{mk}) = \alpha_1 \left(\frac{1}{n} \sum_{p=1}^{n} d_p\right) + \alpha_2(\min\{d_p,\ p = 1,\ n\}), \tag{1}$$

where S represents the set of views; N_{mk} is the number of markers; α_1 and α_2 are user-chosen weighting factors; n is the number of all possible pairs of views; and d_p is the distance measure between the pth pair of views, which is evaluated by an average-pointwise-distance measure of their AOS signatures.

3.1.2. The Optimization Problem

The optimization problem is formulated herein as a two-level hierarchy, where at the upper level the number of markers is minimized. At the lower level, the objective is to select the best set of views, which will maximize the distinctiveness measure defined in Eq. (1):

$$\begin{aligned} &\text{Min} \quad N_{mk} \\ &\qquad \text{Max} \quad F(S, N_{mk}), \quad \text{with } S \subset X \\ &\text{Subject to} \begin{cases} d_p > D_0 \quad p = 1,\ n \\ |S \cap C_{ij}| \geq 2,\ j = 1,\ N_{st};\ i = 1,\ N_{obj} \\ N_{mk} < K_m \end{cases} \end{aligned} \tag{2}$$

where X is the given set of all candidate views, D_0 is a user-chosen threshold for the minimum acceptable distance between a pair of views, C_{ij} is the known set of feasible views for the jth resting-position of object i ($C_{ij} \subset X$), $N_{st}(i)$ is the number of resting-positions of object i, N_{obj} is the total number of objects, and K_m is a given upper bound on the number of markers. The reason for the second constraint, which requires two feasible views for each of the resting positions of all the objects, is to facilitate the multiple-viewpoint recognition strategy.

3.2. A Dynamic-Programming-Based Solution Method

3.2.1. Eliminating Non-suitable Candidate Views

All the flat surfaces of the given objects can be considered as candidates for marking. However, some may not be suitable for marking mainly due to visibility concerns. For this purpose, a set of heuristic rules were defined. These rules state that a candidate surface should be:

- not self-obscured;
- sufficiently large to accommodate a marker;
- the largest amongst several parallel surfaces; and,
- larger than the usual acceptable surface size in a situation where the surface forms a concave geometry with its surroundings.

3.2.2. The Search Technique

The optimization problem formulated in Eq. (2) can be stated equivalently as: Given the total set of views X, and a number of subsets of X corresponding to each of the resting-positions of the objects (namely, C_k, $k = 1$ to N_s, where $N_s = \sum_{i=1}^{N_{obj}} N_{st}(i)$), *determine* the minimum size subset of X, S, such that: (1) for each of the sets C_k ($k = 1, N_s$), at least two of its members are in the set S (i.e., $|S \cap C_k| \geq 2$), (2) the distance between any two members of S is greater than D_0, and (3) the distinctiveness of S (namely, F) is the maximum, if there exist other same-size subsets of X that also satisfy (1) and (2).

Dynamic-programming techniques[21] can be utilized to resolve the optimization expressed in Eq. (2) by treating it as a multi-stage decision-making problem. In dynamic-programming, a decision made at any stage is affected by its predecessors and invariably affects its successors. Namely, employing different sequences would result in different solutions. However, for the problem represented in Eq. (2), there is no given preference concerning the sequence of the stage (i.e., the resting positions). Hence, a global-optimal solution could only be the one that results from considering all the possible sequences of stages.

The stages in this problem can be divided into different groups, where in each group the stages belong to the same object. All the *significant* sequences of stages are considered by generating a two-level permutation. The inner level permutes the stages within each group, and the outer level permutes the groups.

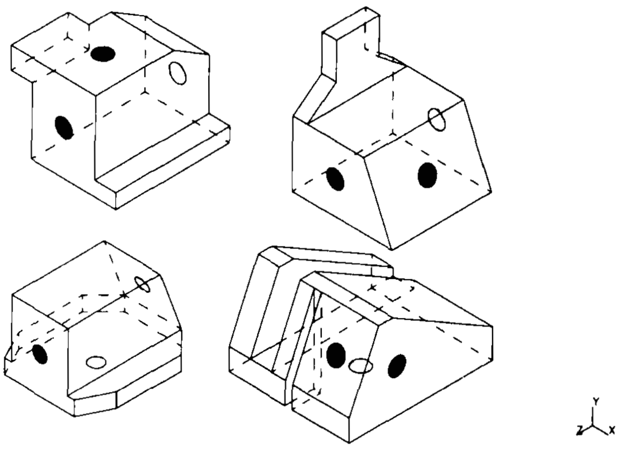

Figure 2. Optimal locations of the markers.

3.3. An Example

A set of four objects having 53 flat surfaces (and 22 resting positions) was considered. Twenty-three candidate views were obtained after applying the heuristic rules. The weighting factors in Eq. (1) were chosen as: $\alpha_1 = \alpha_2 = 0.5$. The optimization procedure yielded the minimum acceptable number of markers as: $N^*_{mk} = 12$, with their best placement on the objects, as shown in Figure 2, corresponding to the highest achievable distinctiveness for 12 markers.

4. OPTIMAL CAMERA PLACEMENT

The detection of a marker is needed under two circumstances: At least one marker should be detected at the beginning of the recognition process, for the acquisition of the first standard-view. An additional marker should be detected to allow the acquisition of a second standard-view, when the first standard-view is insufficient to permit recognition of the object. The former is studied in this section. The latter is addressed in section 6.

4.1. Problem Formulation

Due to the minimization of number of markers, the camera may have to be placed at several locations before the first marker is detected. The problem addressed here is to determine the optimal *initial locations* of the camera to satisfy a desired degree of certainty in detecting at least one marker.

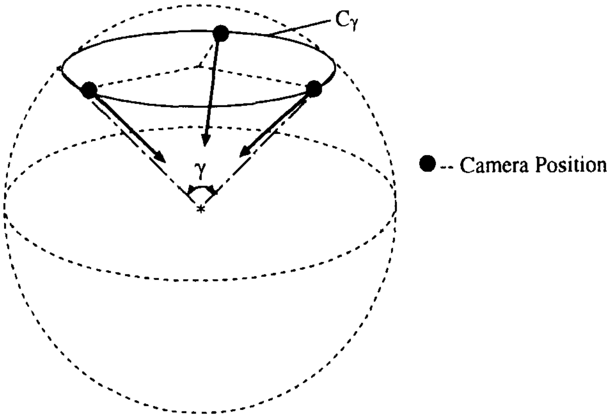

Figure 3. Camera's location-sphere.

4.1.1. Camera's Location Sphere

The mobile camera is assumed to be located on the surface of a virtual sphere, referred to as the camera's *location-sphere* (Fig. 3). For an object observed in the scene and centered with respect to the location sphere of the camera, its rotational orientation about the normal of the support-plane is completely random. To facilitate the calculations, this randomness is treated here in the reverse way:

> *The orientation of the viewed object is treated as if fixed, and the mobile camera is defined to be randomly located on a cross-sectional circle on the location-sphere surface.*

The center-angle of the sphere, γ, which also defines an orientation circle C_γ, is the only variable that specifies the orientation of the camera. For any γ, a number of uniformly spaced positions (N_{cm}) are defined for the camera on C_γ. It is assumed that when the camera does not detect any marker at its initially randomly placed position on C_γ, it can be moved to another position on C_γ. By placing the camera at a maximum of N_{cm} positions on C_γ, the desired degree of certainty in marker detection is achieved. As will be discussed later, the overall objective is to minimize the number of camera movements by minimizing the necessary N_{cm}.

4.1.2. Marker Visibility

Each marker on an object is associated with a 3-D semi-infinite volume, referred to as the *visible-space* of the marker. The boundaries of the visible-space of a marker are determined based on the location of the marker on the object and the object's geometry. In our work, a marker is positioned at the geometric center of a surface to simplify the procedure of marking the workpieces.

The intersections of the visible spaces of all the markers on an object and a specific orientation circle C_γ can be geometrically obtained. These intersections divide C_γ into a finite set of sections (arcs) classified herein as: (1) "detectable-arcs," from which one or more markers are detectable; and (2) "undetectable-arcs," from which no marker is detectable. As explained in a previous work,[19] the classification procedure for a C_γ is a function of N_{cm} as well.

The detectable-arcs can be further divided into two different classes: (1) those from which only one marker can be detected, and (2) those from which two or more markers can be detected. In the latter case, the "goodness" of marker detectability is improved by either knowing directly the location of an additional marker should a second standard-view be required, or having the flexibility of selecting one marker from the several markers detected.

4.1.3. Probability of Detection

Let VS denote the set of detectable-arcs and IVS denote the set of undetectable-arcs, then:

$$C_\gamma = VS \cup IVS. \tag{3}$$

Each member of VS (namely, a specific segment of C_γ) is associated with information that states which of the markers (specified by its identity number) can be observed at which of the camera positions (specified by the sequence number of the camera position).

For each resting-position of an object, the probability of detecting at least one marker is defined as:

$$g(\gamma, N_{cm}) = \frac{1}{2\pi} \int_0^{2\pi} \delta_i d\theta, \tag{4}$$

where the integrand δ_i is defined as:

$$\delta_i = \begin{cases} 1, & \theta \in VS \\ 0, & \theta \in IVS \end{cases}. \tag{5}$$

The variables (γ, N_{cm}) define VS and IVS. It should be remembered that herein object orientational randomness is treated in terms of random placement of the camera on C_γ.

Given the occurrence-probability of each resting-position of an object, p_{ij}, and the occurrence-frequency of each object, $p_n(i)$, the probability of detecting at least one marker with respect to *all* the given objects is then defined as:

$$G(\gamma, N_{cm}) = \sum_{i=1}^{N_{obj}} p_n(i) \left[\sum_{j=1}^{N_{st}(i)} p_{ij} \times g(\gamma, N_{cm}) \right]. \quad (6)$$

Eq. (6) can be modified for considering the visibility of multiple markers only:

$$H(\gamma, N_{cm}) = \sum_{i=1}^{N_{obj}} p_n(i) \left[\sum_{j=1}^{N_{st}(i)} p_{ij} g'(\gamma, N_{cm}) \right], \quad (7)$$

where

$$g'(\gamma, N_{cm}) = \frac{1}{2\pi} \int_0^{2\pi} \delta_j I d\theta. \quad (8)$$

The integrand δ_j is defined as follows:

$$\delta_i = \begin{cases} 1, & \theta \in vsm \\ 0, & \theta \in vss \end{cases}, \quad (9)$$

where *vsm* is the subset of *VS* from which multiple markers are detectable and *vss* is the subset of *VS* from which only one marker is detectable. The integrand I is equal to the number of camera–marker pairs. Each pair represents the detection of one marker from one camera position.

As an overall criterion, the detectability of markers is defined as the combination of the two measures described in Eqs. (6) and (7):

$$J_{dt} = G + \beta H, \quad (10)$$

where β is a positive weighting factor. In Eq. (10), a value of $\beta = 0$ will imply that no special consideration is given to the fact that multiple markers are viewed from certain section of C_γ for non-zero H. On the other hand, as the β value increases, it implies that observing multiple markers is preferable to seeing only one marker.

4.1.4. The Optimization Problem

Based on the previous two sections, one can note that by moving the camera to a finite number of locations on an orientation circle C_γ, the detectability of the "first" marker can be guaranteed. However, the objective of a typical active-vision system would be to minimize the recognition effort. In this context, the optimization problem formulated below targets primarily the minimization of N_{cm}. The detectability of markers is also addressed by maximizing J_{dt} for any given N_{cm}. The two criteria are treated hierarchically in a two-level optimization where the number of camera positions is minimized first:

$$\text{Min} \quad N_{cm}$$

$$\text{Max} \quad J_{dt}(\gamma, N_{cm}), \quad \text{for the specific } N_{cm}$$

$$\text{Subject to} \begin{cases} \gamma \in \Gamma, \text{ where } \Gamma = \{\gamma_i | G(\gamma_i, N_{cm}) \geq p^*\} \\ N_{cm} < K_c \\ 0° \leq \gamma \leq 180° \end{cases} \quad (11)$$

where Γ denotes the feasible range of γ, and K_c is a user-defined upper limit on N_{cm}. When the user-chosen threshold p^* is 1.0, the optimization must yield a set of camera locations specified by (γ, N_{cm}) that will guarantee the detection of at least one marker, namely, $G(\gamma, N_{cm}) = 1$.

4.2. The Solution Algorithm

The optimization expressed in Eq. (11) is solved by a two-level search process. For each N_{cm}, the feasible region Γ of the variable γ is determined to satisfy the constraint imposed on the user-given minimum acceptable probability of detecting at least one marker.

The multi-mode global-extremum optimization algorithm[22] can be used to determine the optimal angle γ^* which yields the maximum J^*_{dt} (with respect to the N^*_{cm}). The angle γ^* defines the desired optimal camera orientation, and the corresponding N^*_{cm} is the desired minimum number of initial camera positions on C_γ.

4.3. An Example

The optimal placement of the initial locations of the camera was carried out for the set of four pre-marked objects shown in Figure 2. For $p^* = 1.0$ (namely, when the probability of detecting at least one marker is 100%), and for $\beta = 1$ in Eq. (10), the minimum number of camera positions was obtained as: $N^*_{cm} = 3$. The best γ angle, which yields the maximum detectability $J_{dt}(\gamma, 3)$, was $\gamma^* = 96°$.

5. OPTIMAL CAMERA-MARKER ARRANGEMENT

The optimal placement of the initial location(s) of the camera was addressed above with respect to a set of objects that were pre-marked. However, one can note that different numbers of markers or different marker-locations on the objects would result in different optimal initial locations of the camera. Therefore, simultaneous solution of the camera placement and object pre-marking problems would be most beneficial for minimizing recognition efforts.

5.1. Problem Formulation

Amongst the four criteria discussed in sections 3 and 4, the most important one is the number of camera positions, because the time spent on moving the camera from one position to another would usually be longer than the time spent in comparing two views. The criterion of secondary importance is the number of markers, because fewer markers implies fewer view comparisons during 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 herein), or optimized hierarchically if one is preferred over the other.

The multi-objective three-level optimization problem is thus defined as:

$$\text{Min} \quad N_{cm}$$
$$\text{Min} \quad N_{mk}$$
$$\text{Max} \quad W(N_{cm}, N_{mk}, S, \gamma) = \lambda_1 J_{dt} + \lambda_2 F$$

Subject to

$$\begin{cases} \gamma \in \Gamma, \text{ where } \Gamma = \{\gamma \mid G(N_{cm}, N_{mk}, S, \gamma) = 1\} \\ d_i > D_0, \; i = 1, n \\ |S \cap C_{ij}| \geq 2, \; d = 1, N_{st}(i); \; i = 1, N_{obj} \\ N_{mk} < K_m \\ N_{cm} < K_c \end{cases} \tag{12}$$

where W is the objective function for the innermost level of the optimization combining the distinctiveness of views and the detectability of markers; J_{dt}, G, Γ, and K_c are as defined previously in section 4 (except that here both G and J_{dt} depend not only on N_{cm} and γ, but also on S and N_{mk}); F, D_0, C_{ij}, and K_m are as defined previously in section 3; and λ_1 and λ_2 are weighting factors.

5.2. Solution Method

To eliminate non-suitable flat-surface candidates, the set of heuristic rules described in section 3.2.1 could be applied here as well. However, after the application of these heuristics, the number of remaining candidates for S would be still quite large. Moreover, the dynamic-programming method developed in section 3 is not quite suitable for the integrated problem addressed here. The reason is that before a complete (sub)set of views for an object is selected, the probability G cannot be calculated for the object, because G should be determined by considering all the markers on the object. This implies that the constraint condition $G = 1$ cannot be examined at every single stage, because a set of views for an object is completely selected only after all the stages corresponding to this object have been reached.

By studying the specific constraints of this optimization problem, an algorithm that significantly reduces the number of feasible combinations was developed. The algorithm is based on the following two properties:

1. Let S_i consist 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 $|S_i \cap C_{ij}| \geq 2$ ($j = 1, N_{st}(i)$) is satisfied for each individual object.
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 1 for any individual object.

The major steps of the algorithm are as follows:

Step 1. Use the solid models of the objects and the set of feasible views for each resting-position, and apply the heuristic rules on all the flat surfaces so that non-suitable surfaces are eliminated.
Step 2. Let $N_{cm} = 1$.
Step 3. For each object, find the minimum number of markers and the corresponding group of candidates of S_i by examining the constraints.
Step 4. Determine the set of feasible candidates of S by combining all groups resulting from Step 3 and examining $d > D_0$ and $G = 1$ for all the possible combinations.
Step 5. If Step 4 results in a non-empty set of S, proceed to Step 8; otherwise, proceed to Step 6.
Step 6. Increase the current value of N_{mk} by 1. If a non-empty set of S has resulted proceed to Step 8. Otherwise, proceed to Step 7.

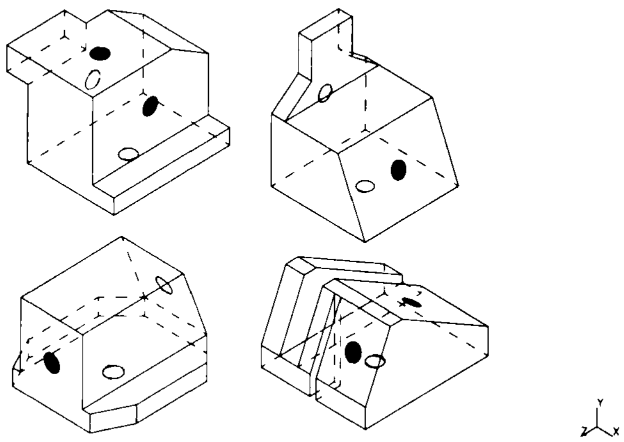

Figure 4. Marker locations resulting from the integrated optimization.

Step 7. If the current N_{mk} is under the given upper limit, return to Step 6 to increase N_{mk} further. Otherwise, increase N_{cm} by 1 if N_{cm} has not reached the given upper limit, and return to Step 3.

Step 8. Determined the best S and γ and stop the procedure.

5.3. An Example

The solution method described in section 5.2 was performed on the same set of four objects shown in Figure 2. For $\lambda_1 = \lambda_2 = 0.5$, the following optimal solution was obtained: $N^*_{cm} = 1$, $N^*_{mk} = 13$, and $\gamma^* = 36°$. The optimal locations of the 13 markers are as shown in Figure 4.

As can be noted, the use of the integrated solution technique reduced the optimal number of camera positions from 3 to 1 (see section 4.3). This was achieved at a minimal trade-off of increasing the number of markers from 12 to 13 (see section 3.3).

5.4. Discussion

In solving the aforementioned optimization problem, several assumptions were made with respect to certain parameters. A sensitivity analysis was hence conducted to study the effect of inaccuracies in such parameters on the objective function W defined by Eq. (12) in section 5.1. The parameters examined were: the occurrence-probability of each resting-position, position of the center of the camera's location-sphere in a pre-defined reference frame, position of the object in the reference frame, the radius of the camera's location-sphere, and the position of the marker's center on a surface (whose nominal value is at the geometric center of a surface, as mentioned in section 4.1.1).

The simulation results[19] showed that the optimization is not very sensitive to uncertainties in these parameters (e.g., if the uncertainty in each of the parameters is under 12%, their *total* effect is less than 5%).

6. SECOND-VIEWPOINT SELECTION

The principal reasons why the first standard-view image could be insufficient to identify a viewed object include:

- The existence of similar standard-views in the database,
- Defective objects, or physical disturbances within the field-of-view of the camera, and
- Misalignment of the optical axis of the camera; a situation that causes a large distortion of the standard-view image.

In regard to the first reason, one can easily know whether there are similar standard-views in the database once the objects are pre-marked. As for the other two reasons, they involve unpredictable events. Thus, the problem to be solved is how to select the next viewing location of the camera so that a new marker can be detected and hence used for the acquisition of the second standard-view. The camera may have to move to several new locations during the search for the second marker, because the whereabouts of a new marker on the viewed object is virtually unknown *a priori*. Those successive locations are referred to as the camera's *secondary locations*. A primary objective in the process of determining these locations is to minimize their number.

It should be noted however, that the objects are assumed to be marked already and the initial camera locations to be determined already. Two methods were developed to address this problem.[19] They are briefly summarized in the next subsection.

6.1. The Probabilistic Solution Method

Based on the marker-arrangements, one can determine the probability of detecting two markers for a given set of pre-marked objects. By determining the corresponding locations of the mobile camera that guarantee the detection of at least two markers, one is assured of detecting a new marker that is different from the one used for the acquisition of the first (standard-view) image.

Concepts introduced in section 4, such as the visible-space of the marker, the location-sphere of the camera, the representation of the location of the camera, and the detectability of the markers, apply here as well. The difference is that the camera locations to be determined here are those that will guarantee the detection of a second marker rather than only an initial marker. These locations of the camera are denoted as γ_2 and N_2.

6.2. The Deterministic Solution Method

In this method, the set of candidate objects consists of those few objects that have similar standard-views. For this situation, the result of comparing the first acquired image with the standard-view database would indicate that there are several standard-views similar to the acquired view. By choosing one of these standard-views, the identity of the viewed object is hypothesized. Then, the known relative locations of the markers on the hypothetically identified object can be used to select a new visible marker and obtain its location.

6.3. A Global Strategy for Determining the Camera's Secondary Location

Based on *a priori* knowledge and run-time information obtained from the first acquired image, one may conclude that the cause for requiring a second image is the existence of similar standard-views in the database. In this case, one can either use the probabilistic method (where the best N_2 and γ_2 are determined with respect to only the few objects that have similar standard-views), or the deterministic method. For other cases, namely where one cannot be sure what the cause is, one should use the probabilistic method (where the best N_2 and γ_2 are determined with respect to all the objects at hand).

7. IMPLEMENTATION

To verify the proposed (on-line) 3-D recognition method, as well as to test the results of the off-line planning procedures developed in this work, an experimental setup of the active-vision system was developed in the CIMLab.

7.1. The Experimental Setup

7.1.1. Hardware Description

The hardware for the on-line recognition process included:

1. *The host platform:* An IBM-compatible PC-486.
2. *The imaging subsystem:* A Hitachi CCD camera, a red-signal filter, a plug-in PIP Matrox digitizer board (640 × 480 resolution), and proper scene illumination.
3. *The robot subsystem:* A GMF S-100 robot and a GMF Karel robot controller.

7.1.2. Software Description

The modules of the on-line recognition process were implemented in the C language on the host as a single sequential program, referred to as the *3-D object recognizer*. The functions of the 3-D object recognizer consisted of:

- detecting the presence of a marker in the scene;
- estimating the 3-D location of a marker, namely estimating a standard-view-axis;
- sending commands to the robot controller concerning the designated camera locations, to request corresponding robot movements;
- analyzing the acquired standard-view image by determining the contour of the image, determining the characteristic ellipse (CE) of the image, and encoding the image with the AOS signature;
- matching an acquired image with the reference standard-views in the database based on the comparison of their AOS signatures;
- searching for an additional marker in the scene when necessary; and,
- estimating the 3-D location of the identified object.

7.2. CAD-based Planning

Research on CAD-based vision has concentrated in the past primarily on the use of a CAD system to generate (1) the representations of objects and (2) recognition strategies.[23–25] In this work, a commercial CAD package—I-DEAS[b]—was employed to cre-

[b] I-DEAS was developed by the Structural Dynamics Research Corporation (SDRC), Milford, Ohio.

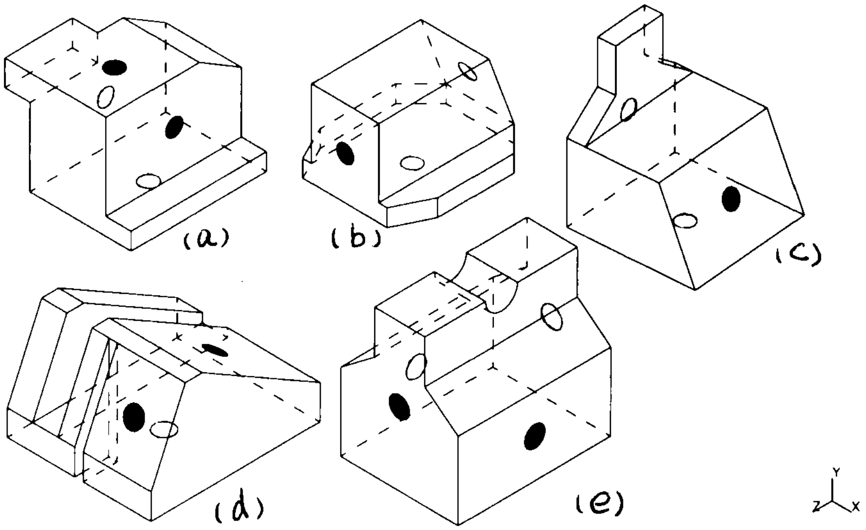

Figure 5. A set of pre-marked objects.

ate 3-D models of the objects under consideration. Synthetic images of candidate views were then generated as *shaded-images* by using the display function of the Solid-Modeling module of I-DEAS. An interface program was developed to seize a shaded-image from the display screen, convert it into a binary image, and store the image file for later processing. Based on the boundary of the binary image, the AOS signature of the view was generated. It should be noted that standard-views generated on a CAD system correspond to "nominal" object representations. Thus, the corresponding AOS signatures form the best noise-free database for future matching purposes. An interface program was also developed to access the object models created by I-DEAS, and to generate the visible-spaces of the markers.

7.3. Experimental Verification

The recognition process was tested with the set of five objects shown in Figure 5. It can be noted that, within the selected set, two objects have similar standard-views. The CAD-based planning stage, which utilized the optimal camera-marker arrangement technique presented in section 5, yielded $N^*_{cm} = 1$, $\gamma^* = 36°$, and $N^*_{mk} = 17$, with the locations of the markers as shown in Figure 5.

7.3.1. Identifying an Object

Experiments were initiated by locating the camera at its optimal initial location (namely, $N_{cm} = 1$, $\gamma = 36°$). For example, with Object-*b* in the scene, two markers were detected initially (Fig. 6). A standard-view was then acquired by selecting one of the markers (Fig. 7). The comparison of the acquired image and the reference standard-views identified Object-*b* correctly with the standard-view-7 as the best match, with the lowest dissimilarity measure of $dis_7 = 0.36$.

The AOS signatures of all the standard-views of all the objects were cross-checked. The results are listed in Table I, where sv_i stands for the *i*th standard-view in the database, and mk_i stands for the standard-view image acquired based on the *i*th marker. The sign "–" indicates that the comparison of the acquired image and a reference view was ter-

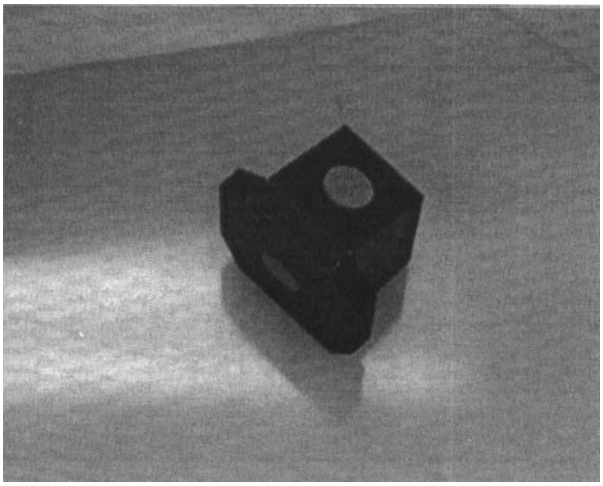

Figure 6. Object-*b* in the scene.

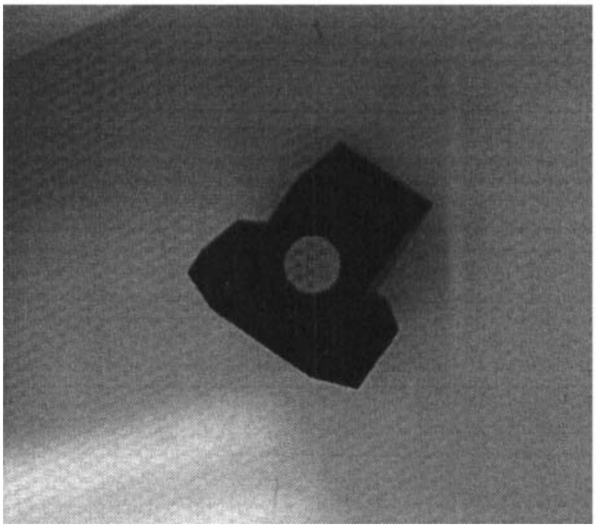

Figure 7. An acquired standard-view image.

minated because the difference between their eccentricities was larger than a given threshold, and hence the reference view was considered to be an automatic "no-match" to the acquired image.

7.3.2. Acquiring Secondary Standard-views

The diagonal elements on Table I show that the identification of the viewed objects was very successful, except for the acquired images corresponding to mk_9 and mk_{14}. This outcome is in agreement with the objective of the (off-line) optimization, which aims at maximizing the distinctiveness of the standard-views. As expected, for the cases of mk_9 and mk_{14}, both sv_9 and sv_{14} (namely, the two similar standard-views) were very similar to each of the acquired images, which implies the need for a second-view.

Numerical results for the secondary camera-location optimization are shown in Table II.

The overall strategy for determining secondary camera locations was thoroughly checked in a previous work.[19] In this article, only an exemplary case is discussed.

When Object-*c* was considered, mk_9 was detected initially (Fig. 8). After the camera was aligned with the standard-view-axis specified by this marker, the first image was acquired. The comparison of this image with the reference views indicated that two of the views were very close to the acquired image (dis_9 = 0.060 and dis_{14} = 0.060). The camera was retained on $C_{36°}$ (because $\gamma_1 = 36° \in [12°, 42°]$) and moved to a different secondary position. A new

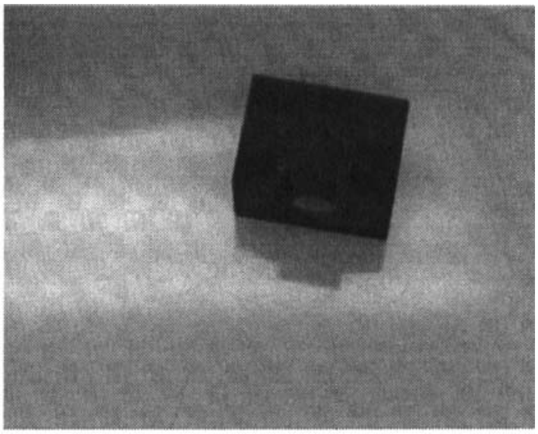

Figure 8. Object-*c* in the scene.

marker (mk_8) was then detected (Fig. 9). The second standard-view image was acquired correspondingly (Fig. 10). This image was then identified as standard-view #8 (dis_8 = 0.165), and the object was correctly identified as Object-*c*.

When Object-*e* was considered, mk_{14} was detected initially. A procedure similar to that described for Object-*c* yielded the successful recognition of this object as well.

8. DISCUSSION AND CONCLUSIONS

Optimal solutions to object pre-marking significantly facilitate the marker-detection and the standard-view matching processes, by providing the minimum number of markers and their optimal locations on the objects. An arbitrarily determined arrangement of markers, on the other hand, would

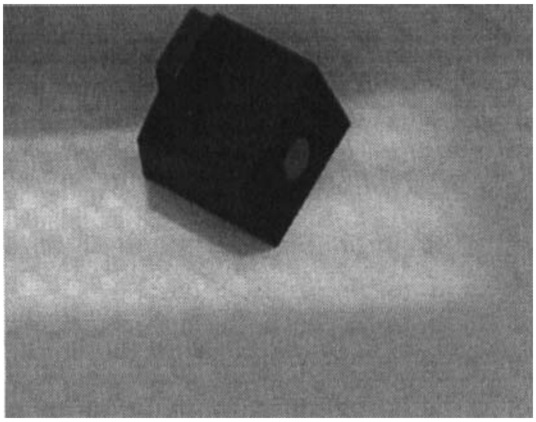

Figure 9. Secondary viewing location for object-*c*.

Table I. Experimental results of standard-view matching.

	sv_1	sv_2	sv_3	sv_4	sv_5	sv_6	sv_7	sv_8	sv_9	sv_{10}	sv_{11}	sv_{12}	sv_{13}	sv_{14}	sv_{15}
mk_1	0.01	10.4	6.9	–	–	18.1	–	10.3	18.9	–	9.6	–	22.2	19.3	–
mk_2	9.97	0.06	17.9	11.1	18.3	13.3	13.6	–	19.4	–	17.8	13.2	21.1	18.9	11.5
mk_3	7.16	18.8	0.10	7.02	16.7	8.86	14.7	–	20.2	–	10.4	7.6	18.3	20.2	–
mk_4	4.22	11.6	6.79	0.03	–	–	16.6	9.82	17.7	–	8.88	–	24.3	19.9	–
mk_5	–	19.6	15.6	–	0.21	26.0	11.4	–	–	24.9	34.9	19.8	15.7	–	27.0
mk_6	–	12.8	9.25	–	27.6	0.09	23.9	–	20.1	–	16.6	5.06	20.7	21.6	5.82
mk_7	–	13.3	13.6	–	8.71	24.6	0.36	–	11.2	32.3	23.1	25.2	32.4	9.01	30.8
mk_8	10.5	–	–	9.59	–	–	–	0.24	–	–	–	–	–	7.25	–
mk_9	16.8	21.0	20.5	18.6	–	20.8	9.50	–	0.07	–	35.0	28.1	29.0	0.08	–
mk_{10}	–	–	–	–	25.8	4.47	–	–	–	0.02	–	2.81	–	–	3.56
mk_{11}	9.05	20.1	9.67	10.5	–	18.9	21.9	–	32.3	–	0.21	20.8	36.7	36.9	–
mk_{12}	–	14.4	7.13	–	24.0	3.32	28.1	–	30.2	–	16.7	0.73	14.6	28.8	3.49
mk_{13}	21.6	20.2	14.9	18.6	23.4	18.4	34.4	–	28.8	–	37.3	11.0	0.11	30.9	–
mk_{14}	19.5	19.0	20.1	19.7	–	21.7	8.81	7.11	0.12	–	37.1	29.0	31.3	0.11	–
mk_{15}	–	12.0	–	–	26.9	5.61	30.9	–	–	3.22	–	3.69	–	–	0.16

Table II. Optimization Results.

	optimal N	optimal γ	feasible γ_2 region
w.r.t. all five objects	$N_2^a = 3$	$\gamma_2^a = 34°$	$[14°, 34°]$
w.r.t. objects c and e	$N_2^s = 3$	$\gamma_2^s = 42°$	$[12°, 42°]$
initial placement	$N_1 = 1$	$\gamma_1^* = 36°$	–

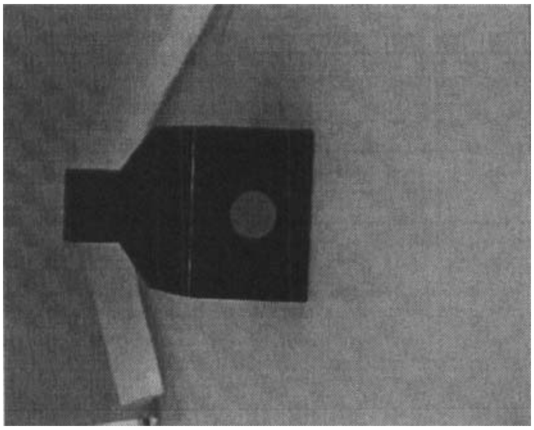

Figure 10. Second standard-view image for object-*c*.

suffer from serious drawbacks such as: (1) ambiguous standard-views resulting from unguided choices of improper marker locations; (2) a greater than necessary number of markers; and, (3) the necessity of choosing one solution (most likely a non-optimal solution) from a very large number of possible marking arrangements (for example, for the four objects discussed in section 3, there would be 3.5×10^{10} possible choices, just by satisfying the constraint of having two markers per resting-position). Furthermore, the impact of proper marker arrangement on camera placement is significant. As shown by an example, the number of camera positions was reduced from 3 to 1 by properly selecting the number and locations of the markers.

The optimal solution for the camera-placement problem on its own provides a cluster of viewing locations for the camera that will guarantee the detection of special features (namely markers) on any viewed object. Without an optimization, one would not know whether the detection of a marker on a randomly oriented object in the scene would be successful within a reasonable number of camera/robot moves. Furthermore, to choose a proper orientation angle for the camera would also be a serious task.

The nature of the marker-arrangement problem addressed herein is very similar to the classical *feature-selection* problem in pattern recognition,[26] where a solution is dependent on the existence of: (a) a capability for evaluating the effectiveness of any subset of a given set of features, and (b) an effective strategy for searching for a best subset amongst all the subsets of features. Thus, the proposed solution methods are potentially applicable as well to the general "feature-selection" problem. The techniques developed for solving the optimal camera-placement problem, on the other hand, also have general applications in active vision, because guaranteeing the detection of a marker, in fact, implies an assurance for the detection of a special feature on the object.

REFERENCES

1. R. Bajcsy, "Active perception vs. passive perception," *Proc. IEEE Workshop Comput. Vision*, Michigan, 1985, pp. 55–59.
2. A. Shmuel and M. Merman, "Active vision: 3D from an image sequence," *Proc. Int. Conf. Pattern Recognit.*, Atlantic City, NJ, 1990, pp. 48–54.
3. R. Safaee-Rad, "An active-vision system for 3D-object recognition in robotic assembly workcells, Ph.D. dissertation, Department of Mechanical Engineering, University of Toronto, Toronto, Canada, 1991.
4. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice Hall, Englewood Cliffs, NJ, 1982.
5. C. G. Crawford, "Aspect graphs and robot vision," *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.* 1985, pp. 382–384.
6. M. R. Korn and C. R. Dyer, "3-D multiview object representations for model-based object recognition," *Pattern Recognit.* **20**(1), 91–103, 1987.
7. T. A. Bodziak, "Using UV metal-marking dyes in industrial vision inspection of complex assemblies," *Proc. Vision '86*, Detroit, MI, 1986, pp. 9-31–9-39.
8. M. R. Kabuka and A. E. Arenas, "Position verification of a mobile robot using standard pattern," *IEEE J. Rob. Autom.* **3**(6), 505–516, 1987.
9. J. W. Bales and L. K. Barker, "Marking parts to aid robot vision," NASA Technical Paper 1819, April 1981.
10. Y. -C. You, J. -D. Lee, J. -Y. Lee, and C. -H. Chen, "Determining location and orientation of a labelled cylinder using point-pair estimation algorithm," *Proc. IAPR Int. Conf. Pattern Recognit.*, The Hague, Netherlands, 1992, pp. 354–357.
11. A. Cameron and H. Durrant-Whyte, "A Bayesian approach to optimal sensor placement, *Int. J. Rob. Res.* **9**(5), 70–88, 1990.
12. K. Tarabanis and R. Y. Tsai, "Planning viewpoints that simultaneously satisfy several feature detectability constraints for robot vision," *Proc. Int. Conf. Adv. Rob.* Pisa, Italy, 1991, pp. 1410–1415.
13. S. Yi, R. M. Haralick, and L. G. Shapiro, "Automatic sensor and light source positioning for machine vision," *Proc. Int. Conf. Pattern Recognit.*, Atlantic City, NJ, 1990, pp. 55–59.
14. H. -S. Kim, "Active multiple view object recognition," Ph.D. dissertation, The University of Michigan, Ann Arbor, MI, 1988.
15. C. I. Connolly, "The determination of next best views," *Proc. IEEE Int. Conf. Rob. Autom.* St. Louis, MO, 1985, pp. 432–435.
16. R. Safaee-Rad, K. C. Smith, B. Benhabib, and I. Tchoukanov, "3D-location estimation of circular fea-

tures for machine vision," *IEEE Trans. Rob. Autom.* **8**(5), 624–640, 1992.

17. R. Safaee-Rad, K. C. Smith, B. Benhabib, and I. Tchoukanov, "Constraints on quadratic curves under perspective projection," *J. Image Vision Comput.*, **10**(8), 532–548, 1992.
18. R. Safaee-Rad, K. C. Smith, B. Benhabib, and I. Tchoukanov, "Application of moments and Fourier descriptors to the accurate estimation of elliptical shape parameters," *J. Pattern Recognit. Lett.* **13**, 497–508, July 1992.
19. X. He, "CAD-based off-line planning for an active-vision system," Ph.D. Dissertation, Department of Mechanical Engineering, University of Toronto, Toronto, Canada, 1993.
20. I. Tchoukanov, R. Safaee-Rad, K. C. Smith, and B. Benhabib, "The angle-of-sight signature for 2D shape analysis of manufactured objects," *Pattern Recog.*, **25**(11), 1289–1305, 1992.
21. L. Cooper and M. W. Cooper, *Introduction to Dynamic Programming*, Pergamon Press, L Oxford, UK, 1981.
22. A. N. Babii, "Algorithm for finding the value of the global extremum of a function of several variables with a given accuracy, *Cybernetics*, **14**(5), 699–703, 1978.
23. B. Bhanu and C. C. Ho, "CAD-based 3D object representation for robot vision," *Computers*, **20**(8), 19–36, 1987.
24. S. J. Dikinson, A. P. Pentland, and A. Rosenfeld. "From volumes to views: An approach to 3-D object recognition," *Proc. IEEE Workshop Directions Autom. CAD-Based Vision*, Maui, Hawaii, 1991, pp. 85–96.
25. K. Ikeuchi, "Generating an interpretation tree from a CAD model for 3D-object recognition in bin-picking task," *Int. J. Comput. Vision*, 145–165, 1987.
26. M. Ichino and J. Sklansky, "Feature selection for linear classifiers," *Proc. Int. Conf. Pattern Recognit.* Montreal, Canada, 1984, pp. 124–127.