New Sensing Strategies for Monitoring Moving Polyhedral Objects by Machine Vision

JIN JANG LEOU AND WEN HSIANG TSAI, MEMBER, IEEE

Abstract **—A new set of sensing strategies for monitoring three-dimensional (3-D) moving objects by computer vision is proposed. Here 3-D object surface points are selected as the features for monitoring 3-D moving objects because the point features are easy to detect, extract, store, and manipulate. It is proved that the minimum measurable feature point set for monitoring a 3-D moving convex polyhedral object is just the set containing all the junction points of the object. Based on the sampling theorem and several properties of photogrammetry, it is proved that the minimum data acquisition rate of a vision system monitoring 3-D moving objects can be determined with discretely sampled two-dimensional image sequence data only. Certain properties of orthographic projection useful for determining the minimum number N_s of sensors needed to monitor 3-D moving convex polyhedral objects are investigated, and the bounds on N_s are also derived. Finally, an algorithm for determining N_s and the corresponding directions of the sensors is proposed. The feasibility of the proposed algorithm is shown by three illustrative examples and an application example.**

I. Introduction

This correspondence is concerned with the application of computer vision to moving object monitoring in three-dimensional (3-D) industrial environments. Computer vision provides a powerful sensory tool for many robot control and industrial automation applications. In a controllable industrial automation environment, there are many operating robots, conveyors, moving parts, storage bins, pallets, fixed obstacles, human workers, etc. The interaction among these "objects" in general is unpredictable, though the local interaction between two adjacent objects may be known. Therefore, we need a high-level machine vision system to monitor the environment.

The main functions of a monitoring vision system include: 1) checking if the robots are in proper operations; 2) checking if the robots have reached their desired positions; 3) preventing collisions among robots, human workers, moving parts, obstacles, etc., and 4) performing fire monitoring. A major step of the monitoring system is to sense interesting moving objects in the environment. Development of effective sensing strategies for monitoring 3-D moving objects by machine vision is the major interest of this study.

Many machine vision systems have been developed in the past two decades. Schmitt *et al.* [1] presented a robot vision system based on two-dimensional (2-D) object-oriented models. The system includes an object model representation, a robot-vision interface, a robot controller, and an integrated user programming environment. Clocksin *et al.* [2] described a model-based visual feedback system for robot arc welding of thin sheet steel to improve the robot positioning accuracy. Ambler *et al* [3] described a versatile computer-controlled system for assembly. Perkins [4] provided a model-based vision system for inspecting flat industrial parts. Haass [5] developed an automatic visual surveillance system for industrial workroom environments with emphasis on the prevention of collisions between an industrial robot and human workers. The system is based on detection, recognition, and tracking of moving objects from digital workroom images. Dodd and Rossol [6] described several vision systems for industrial automation applications, such as a vision-controlled robot system for transferring parts on belt conveyors, an industrial eye that recognizes hole positions in a water pump testing process, etc. Aleksander [7] provided a vision system for industrial applications to improve the accuracy of manufacturing and the quality of industrial products. Miller [29] described a sensor-based control system for robot manipulators. The system uses video image feedback to intercept and track objects moving on a flat conveyor. Yeh *et al.* [33] developed a robot vision system capable of detecting intruders and abnormality in a gantry robot work space to ensure the safe operation of the robot. Most of the foregoing systems are special purposed, 2-D in nature, and unsuitable for monitoring moving objects in 3-D environments.

Chang and Tsai [8] provided a 3-D object inspection system by the use of multiple camera views. Luh and Klaasen [9] developed a 3-D vision system for collision-free robot operations. Hasegawa [10] developed an interactive system for modeling and monitoring a manipulation environment. The system provides a laser pointer as a measuring tool and a display monitor for visualizing the internal database and the real world. Shneier *et al.* [11] proposed a high-level robot vision system which includes four processes: the predictive process, the sensory input analyzing process, the matching process, and the descriptive process. The aforementioned systems are either manual or static in nature and are also unsuitable for monitoring moving objects in 3-D environments.

On the other hand, several problems not treated in the previous approaches will be faced in monitoring moving objects in 3-D environments. The first is as follows: what is the minimum *measurable* feature set for monitoring a moving polyhedral object completely? Hall *et al.* [21] developed a method for selecting the "best" subset of a scene. The solution is shown to be a function of the entire scene, which is hard to deal with. In this study the junction points of polyhedral objects are selected as features because they are easy to detect, extract [12], [13], store, and manipulate [14] in comparison with the other features, such as lines, curves, etc. Furthermore, in practical applications a monitoring system usually cannot continuously "see" moving objects because it can only obtain 2-D sampled image sequence data in discrete times. Therefore, the second problem is, can a vision system continuously monitor moving polyhedral objects using only discrete 2-D image sequence data? If this is possible, what is the minimum data acquisition rate? The third problem is, what are the bounds on the minimum number of sensors necessary for monitoring moving polyhedral objects in a 3-D environment under different motion conditions? The final problem is how to arrange the positions and the directions of the sensors so that the monitoring work can be performed optimally in a certain sense. Grimson [15] derived a technique for predicting optimal sensing positions and developed strategies for determining which objects, from a set of objects, are consistent with the sensory data. This technique is applicable to static objects only. In this study, based on the convexity properties [16], the sampling theorem [17], and several properties of photogrammetry and projection processes [18], [19], a set of sensing strategies for monitoring moving polyhedral objects in 3-D environments is developed which removes the weakness of the aforementioned systems and answers the above problems satisfactorily.

In this correspondence we make the following assumptions: 1) the moving polyhedral objects are convex and rigid; 2) the sensors are fixed, i.e., their positions and directions are fixed and known with respect to a global or world coordinate system; 3) the coordinate system is Cartesian; 4) the projection from a 3-D object space onto any 2-D sensory image plane is orthographic [19], and 5) the correspondence processes [20] for tracking tokens

Manuscript received November 11, 1987; revised December 22, 1988. This work was supported in part by the National Science Council, Republic of China, under Grant NSC-77-0404-E-009-017.

J. J. Leou is with the Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan 30050, Republic of China.

W. H. Tsai is with the Department of Information Science, National Chiao Tung University, Hsinchu, Taiwan, 30050, Republic of China.

IEEE Log Number 8926810.

0018-9472/89/0700-0872$01.00 ©1989 IEEE

in distinct images are completely solved. Because the monitoring of 3-D moving objects involves 3-D data acquisition and multiple sensor image data, we are basically faced with a multiple image sequence analysis problem in a 3-D space.

In Section II the minimum measurable feature point set for monitoring a moving convex polyhedral object is derived. In Section III, based on the sampling theorem and several properties of photogrammetry, it is proved that the minimum data acquisition rate of a vision system monitoring 3-D moving objects can be determined with discretely sampled 2-D image sequence data only. Derived in Section IV are the upper and the lower bounds of the minimum number of sensors required for monitoring moving convex polyhedral objects in a 3-D environment under different object motion conditions. An algorithm for the determination of the minimum number of sensors and their corresponding directions is proposed. Three illustrative examples and an application example are included in Section V to verify the proposed algorithm. Conclusions are given finally in Section VI.

II. Minimum Measurable Feature Point Set for Monitoring a Moving Convex Polyhedral Object

Ballard and Brown [14] described three general classes of representations for rigid solids, namely, surface or boundary, sweep (in general, general cylinders), and volumetric. For the surface representation it was claimed that the enclosing surface, or boundary, of a well-behaved three-dimensional object should unambiguously specify the object. For a polyhedral object its boundary is just a set of planes which is well behaved. The set of planes can be represented by a set of plane equations or a set of points. Here, the latter is selected as the feature set because 3-D points are easier to handle. In addition, Ballard and Brown [14] also claimed that a sensory system can be used to *completely* monitor a moving polyhedral object if the monitored moving polyhedral object can be unambiguously reconstructed at any time instant. For the monitoring purpose, every feature point must be physically existent and directly measurable by the sensory vision system. Thus our objective is to find a physically existent and directly measurable 3-D feature point set with a minimum number of points, called *the minimum measurable feature point set*, by which the monitored moving object can be unambiguously reconstructed at any time instant. Since the monitored polyhedral object is assumed to be convex, the minimum measurable feature point set can be determined by the use of the convex property [16] of the object, as discussed in the following.

Lemma 1: The convex hull of a set of N points in a 3-D space can be computed in optimal time $O(N \log N)$.

The derivation of the lemma can be found in Preparata and Shamos [22] and is omitted here. A divide-and-conquer algorithm to attain the optimal bound can be found in Preparata and Shamos [22] and also in Preparata and Hong [23].

Lemma 2: The convex hull of all the junction points of a convex polyhedron is just the convex polyhedron itself. Reversely, if the total number of the junction points of a convex polyhedron is N, then the convex polyhedron cannot be uniquely reconstructed with less than N measurable 3-D points or with any other N measurable 3-D points.

Discussions about the foregoing lemma can be found in Rockafellar [16] and the proof is omitted here. The following theorem is an answer to the first problem mentioned previously in Section I.

Theorem 1: The minimum measurable feature point set for monitoring a convex polyhedral object moving in a 3-D space is the set containing all the junction points of the object (i.e., the size of the minimum measurable feature point set is just equal to the total number of the junction points of the object).

Proof: By Lemma 2, if any of the junction points of the convex polyhedral object is dropped, or if they are replaced by any other N measurable 3-D points, then the object cannot be uniquely reconstructed (i.e., the sensory system cannot completely monitor the moving convex polyhedral object). On the contrary, if any other measurable 3-D points are inserted into the junction point set, then the inserted points will contribute no additional information to the sensory system for monitoring the moving object. Furthermore, by Lemma 1 the inserted points will make the reconstruction of the moving object more complicated. This completes the proof of the theorem.

Theorem 1 provides a theoretical basis to obtain the minimum measurable feature point set of a moving convex polyhedral object in a 3-D space when the object is monitored by a sensory system. It says that if the coordinate information of all the junction points of the convex polyhedral object are known at any time instant, the sensory system can monitor the object completely.

By the way, Ikeuchi [34] proposed an extended Gaussian image for 3-D object representation and recognition. By the use of the extended Gaussian image, every face of a convex polyhedron can be represented by an equivalent point specified by three parameters (two angles and a weight). For example, a cube can be uniquely represented by six equivalent points instead of eight junction points by the use of the extended Gaussian image. In practice, however, the equivalent points are not physically existent and not directly measurable by the sensory vision system. Therefore, for the monitoring purpose, the minimum measurable feature point set consisting of the object junction points is used in this study.

III. Determination of Minimum Data Acquisition Rate from Discrete Orthographically Projective Images

For practical applications, the trajectory of any monitored object is assumed to be smooth enough. Thus the trajectory $P_i(t)$ of a specific 3-D object point P_i at time instant t can be described by three position functions $(X_i(t), Y_i(t), Z_i(t))$ in a Cartesian coordinate system, where $X_i(t)$, $Y_i(t)$, and $Z_i(t)$ are all continuous bandlimited signals. In this study, $X_i(t)$, $Y_i(t)$, and $Z_i(t)$ are treated as three orthogonal continuous band-limited signals. By Theorem 1, a sensory system can monitor a moving convex polyhedral object completely if the sensory system can obtain the coordinate information of all the junction points of the monitored object at any time instant. In practice, the sensory system cannot continuously "see" the moving object because the available information is just the discretely sampled 2-D image sequence data. In the remainder of this section it will be proved that discretely sampled 2-D image sequence data are sufficient for monitoring 3-D moving objects if the data acquisition rate is larger than a specific value. This answers the second problem mentioned in Section I.

Theorem 2 (One-Dimensional Sampling Theorem): Consider a continuous band-limited signal $s(t)$ with the maximum frequency component f_M. If $s(t)$ is sampled periodically with sampling frequency $f_s \geqslant 2 \cdot f_M$, then $s(t)$ can be completely reconstructed with discretely sampled data only.

The proof of Theorem 2 can be found in [17] and is omitted here. The sampling theorem says that a continuous band-limited signal can be exactly interpolated at any time instant with discretely sampled data only. The value $f_N = 2 \cdot f_M$ is called the minimum data acquisition rate or the Nyquist rate of the signal.

Lemma 3: Given a continuous composite signal $s(t) = a_0 + a_1 \cdot s_1(t) + a_2 \cdot s_2(t) + \cdots + a_m \cdot s_m(t)$, if $s_1(t), s_2(t), \cdots, s_m(t)$ are m orthogonal continuous band-limited signals, t is a real variable, and $a_0, a_1, \cdots, a_m$ are $(m+1)$ real constants, then $s(t)$ is also a continuous band-limited signal. If $f_1, f_2, \cdots, f_m$ are the

Nyquist rates of $s_1(t), s_2(t), \cdots, s_m(t)$, respectively, then the Nyquist rate f_N of $s(t)$ is equal to the maximum of $f_1, f_2, \cdots, f_m$.

Similar concepts of Lemma 3 can be found in [17], and the proof is omitted here. From Lemma 3 it is concluded that the Nyquist rate of a composite signal is equal to the maximum value of the Nyquist rates of all the composing signals. Let the trajectories of any M object points of a moving convex polyhedral object be represented as $P_i(X_i(t), Y_i(t), Z_i(t))$, and f_{i1}, f_{i2}, f_{i3} be the Nyquist rates of the position functions $X_i(t)$, $Y_i(t)$, and $Z_i(t)$, respectively, $i = 1, 2, \cdots, M$. It follows from Theorem 2 and Lemma 3 that the Nyquist rate f_N of the sensory system monitoring the M object points of a convex polyhedral object is equal to the maximum value of the Nyquist rates of the $3 \cdot M$ position functions of the M object points, i.e., f_N is the maximum of $f_{11}, f_{12}, f_{13}, \cdots, f_{M1}, f_{M2}, f_{M3}$.

In practice, the 3-D continuous position functions of any point of a monitored convex polyhedral object are not directly available. Instead, the only available data are the 2-D projective image sequence data obtained in discrete times. Can the Nyquist rate f_N of the sensory system monitoring the moving object be determined using only the discretely sampled 2-D image sequence? The answer is positive as discussed in the following.

Lemma 4: If the image projection from a 3-D object space onto any 2-D image plane is orthographic, then the 2-D position functions $(U(t), V(t))$ of an object point P on an image plane are the affine transforms of the position functions $(X(t), Y(t), Z(t))$ of P in a 3-D space, i.e.,

$$U = a_1 \cdot X + a_2 \cdot Y + a_3 \cdot Z + a_4 \quad (1)$$

$$V = b_1 \cdot X + b_2 \cdot Y + b_3 \cdot Z + b_4 \quad (2)$$

where $a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$ are real constants, and the coordinate systems of 2-D and 3-D spaces are both Cartesian.

Proof: Let the camera coordinate system and the world coordinate system be related with positional translations (X_c, Y_c, Z_c) and angular rotations (θ, ϕ, ψ), where θ, ϕ, ψ specify the pan, the tilt, and the swing angles, respectively. Additionally, let the deviation of the imaging center in the image plane from the origin of the camera coordinate system be (u_0, v_0), and let the image coordinate scaling factors with respect to the world coordinate system be k_u and k_v. If the homogeneous coordinate representation is used, then by the results in [18], [24] and simple calculations, it can be derived that

$$\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = [\mathrm{INT}][\mathrm{EXT}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (3)$$

where

$$[\mathrm{EXT}] = \begin{bmatrix} a & b & c & p \\ d & e & f & q \\ g & h & i & r \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$[\mathrm{INT}] = \begin{bmatrix} k_u & 0 & 0 & u_0 \\ 0 & 0 & k_v & v_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

and a through i, and p, q, and r are functions of X_c, Y_c, Z_c, θ, ϕ, and ψ only. Thus a through i, and p, q, and r as well as k_u, k_v, u_0, v_0 can be determined when the relationship between the image coordinate system and the world coordinate system is determined. By (3)–(5) we have

$$U = k_u \cdot a \cdot X + k_u \cdot b \cdot Y + k_u \cdot c \cdot Z + (k_u \cdot p + u_0) \quad (6)$$

$$V = k_v \cdot g \cdot X + k_v \cdot h \cdot Y + k_v \cdot i \cdot Z + (k_v \cdot r + v_0) \quad (7)$$

which indicate that the 2-D image position functions (U, V) of an object point are the affine transforms of the position functions (X, Y, Z) of the object point in the 3-D space. This completes the proof of Lemma 4.

Lemma 5: Assume that point $P(X(t), Y(t), Z(t))$ in a 3-D space is orthographically projected onto two distinct image planes with $(U_1(t), V_1(t))$ and $(U_2(t), V_2(t))$ as the resulting point coordinates (position functions) on the two image coordinate systems, respectively. If (by Lemma 4)

$$U_1 = a_{11} \cdot X + a_{21} \cdot Y + a_{31} \cdot Z + a_{41} \quad (8)$$

$$V_1 = b_{11} \cdot X + b_{21} \cdot Y + b_{31} \cdot Z + b_{41} \quad (9)$$

$$U_2 = a_{12} \cdot X + a_{22} \cdot Y + a_{32} \cdot Z + a_{42} \quad (10)$$

$$V_2 = b_{12} \cdot X + b_{22} \cdot Y + b_{32} \cdot Z + b_{42} \quad (11)$$

where the coefficients $a_{11}, a_{21}, \cdots, b_{32}, b_{42}$ are real constants, then

$$a_{11}^2 + b_{11}^2 + a_{12}^2 + b_{12}^2 > 0 \quad (12)$$

$$a_{21}^2 + b_{21}^2 + a_{22}^2 + b_{22}^2 > 0 \quad (13)$$

$$a_{31}^2 + b_{31}^2 + a_{32}^2 + b_{32}^2 > 0. \quad (14)$$

Based on the results in [24], Lemma 4, and some computations, Lemma 5 can be derived easily. The detailed derivation is too long to be included and is omitted here. By Lemma 5 it is observed that any of the three 3-D position functions $(X(t), Y(t), Z(t))$ of an object point must be the contributor of at least one of its 2-D position functions of the corresponding projected points on two distinct image planes (i.e., any of the three 3-D position functions $(X(t), Y(t), Z(t))$ must be the composing function of at least one of the four position functions, $(U_1(t), V_1(t))$ and $(U_2(t), V_2(t))$, of the projected points on two distinct image planes).

Lemma 6: Assuming that a moving object point $P(X(t), Y(t), Z(t))$ in a 3-D space is orthographically projected onto two (or more) distinct image planes, then the Nyquist rate of a sensory system monitoring the 3-D moving object point P is equal to the maximum value of the Nyquist rates of the position functions of the corresponding projected points of P on these distinct 2-D image planes.

The lemma can be derived easily using Lemmas 3 and 5, so the derivation is also omitted here.

Based on the foregoing results, it can be concluded that the Nyquist rate of a sensory system monitoring a 3-D moving convex polyhedral object represented by all its surface points is equal to the maximum value of the Nyquist rates of the position functions of all the projected object surface points on 2-D image planes. However, this conclusion is useless because a real object might consist of an infinite number of surface points. Actually, if a 3-D moving object is represented by all object surface points and if every object surface point P is described by three orthogonal band-limited position functions $(X(t), Y(t), Z(t))$, then by signal theory [30] the Nyquist rate of the sensory system monitoring the moving object can be determined by the object point producing the largest position change rate (velocity). In the following it will be proved that the object surface point of a moving convex polyhedral object producing the largest position change rate must be one of the junction points of the object. Thus the Nyquist rate of the sensory system monitoring the moving object can be determined only by the position functions of all the junction points of the moving object.

Lemma 7: The orthographic projection of a 3-D convex set C on a 2-D subspace L is another convex set.

The derivation of Lemma 7 can be found in [16] and is omitted here.

Lemma 8: For a convex polyhedral object moving in a 3-D space, if every surface point of the moving object is described by three Cartesian position functions, then the object surface point producing the largest position change rate should be one of the junction points of the moving object.

Proof: From Rogers and Adams [31] it is known that any 3-D rigid body motion is equivalent to a rotation by an angle θ around an axis through the origin, followed by a translation $(\Delta X, \Delta Y, \Delta Z)$. For the translation part all the surface points of the convex polyhedral object have an identical amount of positional deviation. Therefore, we only have to focus on the amount of deviation of the rotational part. For those points rotating about an axis through the origin, the larger the distance between an object point and the axis, the larger the positional deviation of the object point within a certain amount of time interval. By Lemma 7 and certain properties of convex polyhedral objects [16], an object point having the largest distance between the point and the rotation axis in a 3-D space or in a 2-D orthographic projection space must be one of the junction points of the moving convex polyhedral object. This completes the proof of this lemma.

The following is the main result of this section.

Theorem 3: If the projection from a 3-D space onto any 2-D image plane is orthographic and if every junction point of a moving convex polyhedral object in the 3-D space is projected onto at least two distinct image planes, then the Nyquist rate of the sensory system monitoring the moving object can be determined using only the projected 2-D position functions of all the junction points of the moving object on distinct image planes.

Proof: By Lemma 8 and signal theory [30], it can be claimed that the Nyquist rate of the sensory system monitoring a moving convex polyhedral object is the maximum value of the Nyquist rates of all the junction points of the moving object in the 3-D space, i.e., the Nyquist rate of the sensory system is determined if the Nyquist rates of all the junction points of the moving object in the 3-D space is determined. By Lemma 6, the Nyquist rate of an object junction point P in the 3-D space is determined if the Nyquist rates of the position functions of the corresponding 2-D projected junction points of P on two (or more) distinct image planes are determined. Since every junction point of the moving object is assumed to be projected onto at least two distinct image planes and since the projection process is assumed to be orthographic, the Nyquist rate of the sensory system is equal to the maximum value of the Nyquist rates of the position functions of all the 2-D projected object junction points. This completes the proof of the theorem.

By Theorem 3, the Nyquist rate of the sensory system is determined if the projected 2-D position functions of all the junction points of the monitored moving convex polyhedral object on distinct image planes are known. However, the "continuous" projected 2-D position functions of all the junction points of the object usually are not directly available. The available data are just "discretely sampled" 2-D image sequence data. This seems to be a problem. Fortunately, in industrial automation applications a moving object usually follows a specific periodic trajectory or a repetitive sequence of primitive motions [25], [26]. Thus a solution can be proposed as follows. First, the sensory system monitors the moving object with a very high data acquisition rate in the "learning" stage. The 2-D position functions of all the projected junction points of the moving object are then reconstructed by the use of the discretely measured 2-D image data based on the sampling theorem. Then the Nyquist rate of the sensory system can be determined according to Theorem 3. Finally, a suitable data acquisition rate (larger than or equal to the Nyquist rate) can be chosen for use in the "operation" stage.

IV. Determination of the Number and the Corresponding Directions of Sensors

When a convex polyhedral object moving in a 3-D space is to be monitored by a sensory system containing several fixed sensors, two important problems as mentioned in Section I should be solved first. The first is, what are the bounds on the minimum number of sensors necessary for monitoring various moving convex polyhedral objects in a 3-D space under different motion conditions? The second is what is the optimal way to arrange the positions and the directions of the sensors? Here by optimality, it is meant that the sensors as a group can completely "see" the minimum measurable feature point set at any time instant. In this section, several properties of orthographic projection are first investigated. Next, the bounds on the minimum number of sensors necessary for monitoring moving objects completely in a 3-D space are derived. Finally, an algorithm for the determination of the minimum number of sensors and their corresponding directions is proposed.

A. Properties of Orthographic Projection and Bounds on the Number of Sensors

An image plane in a camera system can be described by six camera parameters $(X_c, Y_c, Z_c, \theta, \phi, \psi)$ [18], where X_c, Y_c, Z_c are the positional translations, and θ, ϕ, ψ specify the pan, the tilt, and the swing angles, respectively, of the camera with respect to a world coordinate system. The determination of optimal sensor positions and orientations is equivalent to the search of a set of image planes (or their parameters). The previous representation of an image plane is simple, but it is useless in this study because X_c, Y_c, Z_c are three real numbers, so the search space of various image planes is an unbounded space containing an infinite number of search points. Therefore, the range of the search space of various image planes for the determination of sensor numbers and locations has to be treated from another point of view as described in the following.

Definition 1: If an orthographically projective image of a moving convex polyhedral object is always a translational and/or rotational version of another orthographically projective image of the same moving object at any time instant, then these two sensors (image planes) will be treated as an identical one in this study because these two sensors always acquire an identical amount of information about the moving object when the sensory system monitors the object.

For example, as shown in Fig. 1, there are three orthographically projective images of a single object on three image planes at a specific time instant. The projected object image in Fig. 1(b) is a rotational version of that in Fig. 1(a) and the projected object image in Fig. 1(c) is a translational and rotational version of that in Fig. 1(a). Thus these three image planes will be treated as an identical one in this study.

It is assumed that the range of the trajectory of a monitored convex polyhedral object moving in a 3-D space is finite, and that the size of the moving object is also finite. Theoretically, an infinite number of enclosing spheres with different sphere centers and different sphere radii can be selected to enclose the entire range of the object trajectory. Any plane tangent to a specific enclosing sphere corresponds exactly to an orthographically projective image plane. Therefore, the search space of orthographically projective image planes can be transformed to be the tangent planes of all the 3-D enclosing spheres.

Lemma 9: The set containing the tangent planes of all the spheres with different sphere centers and different radii (except

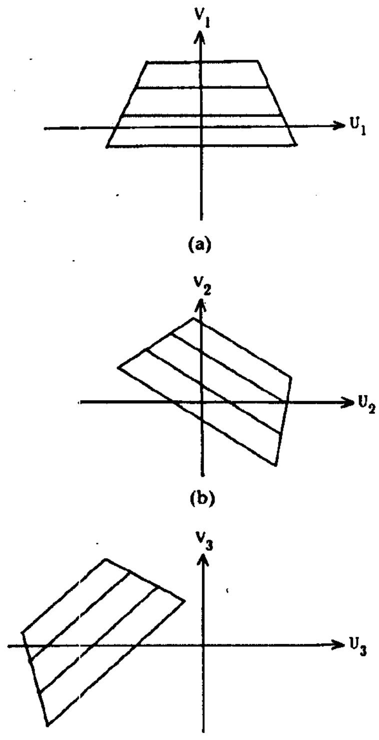


Fig. 1. Three orthographically projective images of a single object on three image planes at specific time instant. (a) Object image on first image plane. (b) Object image on second image plane. (c) Object image on third image plane.

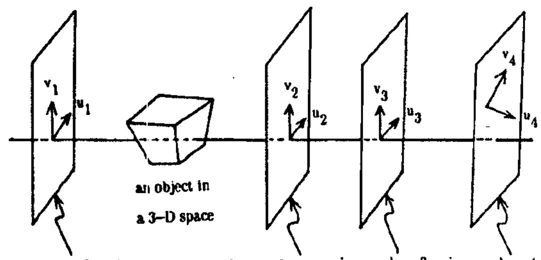


Fig. 2. Four orthographically projective images in a 3-D space, where image planes 2, 3, 4 are mutually parallel with different image coordinate systems.

the degenerate case: a point) is equal to the set containing the tangent planes of all the spheres with a single sphere center but with different radii.

Similar concepts of Lemma 9 can be found in Riddle [32], and the derivation is omitted here. By Lemma 9, the search space of various image planes can be transformed to be the tangent planes of all the enclosing spheres with a single sphere center but with different radii. However, by Definition 1 corresponding tangent planes of all the enclosing spheres with an identical sphere center, but with different radii, will be treated as identical ones. Accordingly, the three corresponding tangent planes (image planes 2, 3, 4) shown in Fig. 2 should be considered as identical. On the contrary, image planes 1 and 2 in Fig. 2 are regarded as distinct because from these two image planes different information about the monitored moving object can be acquired. By the earlier discussions, the search space of orthographically projective image planes (corresponding to sensor positions and directions) can be reduced to include only the tangent planes of a selected enclosing

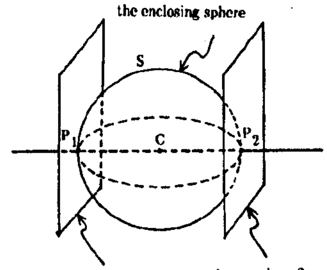


Fig. 3. Spherical tangent image plane pair in 3-D space, where image plane 1 is parallel to image plane 2.

sphere. Such planes can be described by the solid angles of the selected enclosing sphere. A solid angle can be specified by two directional angle parameters, the pan angle θ and the tilt angle ϕ. The ranges of θ and ϕ are $-\pi \leqslant \theta < \pi$ and $-\pi/2 \leqslant \phi < \pi/2$ (or $-\pi/2 \leqslant \theta < \pi/2$ and $-\pi \leqslant \phi < \pi$) with periods 2π and π (or π and 2π), respectively. Thus the sensors can be distinguished just by two parameters, the pan and the tilt angles.

Definition 2: A spherical tangent image plane pair are two distinct parallel tangent planes P_1 and P_2 of an enclosing sphere S of a moving object such that the centroid of S and the two tangent points of P_1 and P_2 to S are collinear as illustrated by the example shown in Fig. 3.

Lemma 10: If the projection is orthographic, then all the junction points of a monitored moving convex polyhedral object over its entire trajectory are visible on each spherical tangent image plane pair of any enclosing sphere, i.e., any junction point of the monitored object can be sensed by at least one plane of the image plane pair at any sampling time instant.

Similar concepts of the foregoing lemma can be found in [16], and the proof is omitted here.

Lemma 11: Either for perspective projection or for orthographic projection, the 3-D coordinate information of a point is recoverable if the point is visible on (or appears on) two (or more) distinct image planes.

The foregoing lemma is the fundamental theory in photogrammetry [18], so the derivation is also omitted here. Assuming that the purpose of object monitoring is to recover the 3-D coordinate information of all the object junction points from 2-D image planes, we have the following results.

Theorem 4: If the projection process is orthographic, then the minimum number N_s of sensors necessary for monitoring a moving convex polyhedral object under various motion conditions is bounded by $2 \leqslant N_s \leqslant 4$.

Proof: By Lemma 11, it is easy to see that $N_s \geqslant 2$. On the other hand, if a sensory system consists of two distinct spherical tangent plane pairs (i.e., the system includes four distinct image planes), then according to Lemma 10 any junction point of the moving convex polyhedral object will be visible on at least two of these four distinct image planes over all its trajectory. By Lemma 11, the 3-D coordinate information of every object junction point is recoverable. This means that the sensory system can monitor the moving object completely. Therefore, $N_s \leqslant 4$. This completes the proof of the theorem.

B. Further Properties of Orthographic Projection

Some properties of orthographic projection are derived here for use in the proposed algorithm for the determination of the minimum number of sensors and their corresponding directions.

Definition 3: A full sensing direction (θ_F, ϕ_F) is the direction of an image plane such that all the junction points of the monitored moving polyhedral object are visible on the image plane over the entire trajectory of the object at any time instant.

Definition 4: The optimal sensing direction (θ_M, ϕ_M) is one in which the image plane can sense the maximum number of object junction points over the entire object trajectory.

For a moving convex polyhedral object there may exist multiple full sensing directions and/or multiple optimal sensing directions.

Lemma 12: For an orthographic projection process, a sensory system can monitor a moving convex polyhedral object completely with only two sensors if the enclosing sphere of the moving object consists of two (or more) distinct full sensing directions.

Lemma 12 can be easily derived from Definition 3 and Lemma 11.

Lemma 13: For a sensory system monitoring a 3-D moving convex polyhedral object under an orthographic projection process, if the optimal sensing direction of the corresponding enclosing sphere is not a full sensing direction, then the minimum number N_s of sensors required for complete monitoring is either 3 or 4.

Lemma 13 can be easily derived from Lemma 11 and Theorem 4, and so the proof is omitted.

Lemma 14: Under the conditions that the projection process is orthographic and that a sensory system uses F sampling periods to monitor completely the entire trajectory of a moving convex polyhedral object with N junction points according to the sampling theorem, if the total number of the junction points of the object visible from the optimal sensing direction of the sensory system over the total F sampling periods is smaller than $2 \cdot F \cdot N/3$, then the minimum number N_s of sensors required for the sensory system is 4.

Proof: Since the sensory system uses F sampling periods to monitor completely the entire trajectory of the moving convex polyhedral object with N junction points, the minimum total number of the object junction points sensed by all the sensors of the sensory system over the F sampling periods is $2 \cdot N \cdot F$ because by Lemma 11 any of the N junction points has to appear on at least two distinct image planes in every sampling period for the monitoring to be complete. If the total number of the object junction points visible on the image plane from the optimal sensing direction over the F sampling periods is smaller than $2 \cdot F \cdot N/3$ and if the number of the sensors N_s of the sensory system is either 3 or 2, then the total number of the object junction points seen on all the image planes of the sensory system over the F sampling periods is smaller than $2 \cdot F \cdot N$, resulting in a contradiction. Therefore, $N_s > 3$ and by Theorem 4, N_s can only be 4. This completes the proof of the lemma.

C. Proposed Algorithm for the Determination of the Minimum Number of Sensors and their Directions

From previous discussions it is seen that the determination of the minimum number of sensors depends only on two parameters, the pan angle θ and the tilt angle ϕ, of the image planes (sensors). The ranges of θ and ϕ are $-\pi \leqslant \theta < \pi$ and $-\pi/2 \leqslant \phi < \pi/2$ with periods 2π and π, respectively. That is, the search space for the determination of N_s is the block area shown in Fig. 4(a), which contains an infinite number of points. In practice, we can quantize this "continuous" search space into a discrete one including $P \cdot Q$ quantized search points, as illustrated in Fig. 4(b). Each quantized point corresponds to a specific image plane.

As mentioned in Section III, the sensory system usually monitors a moving object repeating a certain trajectory or doing a sequence of primitive motions [25], [26] in most automation applications. The sensory system can use many sensors and a sufficiently high data acquisition rate in the "learning" stage, and some parameters of the sensory system, such as the Nyquist rate and the minimum number of sensors of the sensory system, then can be determined for use in the "operation" stage, based on the discrete 2-D measured image sequence data. In the following proposed algorithm the total number of the junction points of the moving object and the minimum number of sensors necessary for the sensory system are assumed to be as N and N_s, respectively.

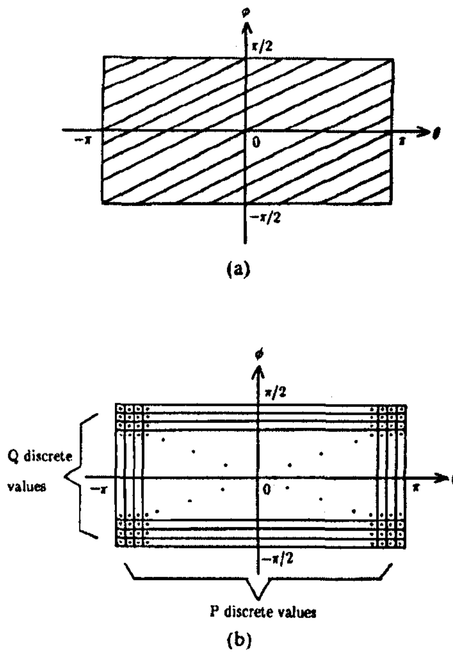


Fig. 4. Search spaces of pan and tilt angles. (a) Continuous search spaces of pan angle θ and tilt angle ϕ. (b) Discrete search spaces of pan angle θ and tilt angle ϕ.

Algorithm: Determination of the Minimum Number of Sensors

Step 1—Initialization: Use four sensors (according to Theorem 4) to monitor the moving convex polyhedral object over its entire trajectory with a sufficiently high data acquisition rate.

Step 2—Nyquist: rate f_N determination: Based on the measured data in Step 1, determine the Nyquist rate f_N of the sensory system (according to Lemma 3) to be the maximum value of the Nyquist rates of the $3 \cdot N$ position functions of the N junction points of the moving object.

Step 3—Resampling: Resample all the trajectories (position functions) of the N junction points of the monitored object with the system Nyquist rate f_N, resulting in totally $4 \cdot F$ sampled image sequence data in the F sampling periods.

Step 4—Object reconstruction and enclosing sphere selection: Reconstruct the moving object over the F sampling periods and select an enclosing sphere which can enclose the object over its entire trajectory.

Step 5—Orthographic projection: Use orthographic projection to project all the junction points of the moving object onto the $P \cdot Q$ distinct image planes (as illustrated in Fig. 4) of the enclosing sphere over the F sampling periods.

Step 6—Counting and indexing: Count and index the total number of junction points visible on every one of the $P \cdot Q$ distinct image planes over the F sampling periods.

Step 7—Full sensing direction detection: If the enclosing sphere contains two or more distinct full sensing directions, then determine the minimum number N_s of sensors (according to Lemma 12) to be two, select arbitrarily two distinct full sensing directions as the sensor directions, and exit; otherwise, go to Step 8.

Step 8—Optimal sensing direction determination: If the total number of the junction points of the monitored object over the F sampling periods visible from the optimal sensing direction is smaller than $2 \cdot F \cdot N/3$, then determine N_s (according to Lemma

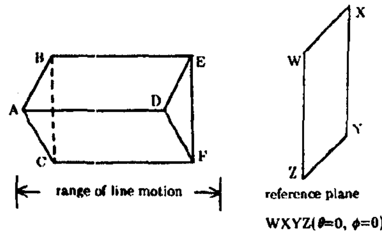


Fig. 5. Determination of minimum number N_s of sensors for simple moving convex polyhedral object in 3-D space.

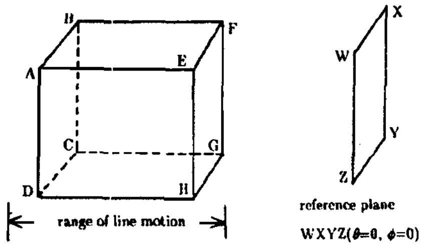


Fig. 6. Determination of minimum number N_s of sensors for moving cube in 3-D space.

14) to be 4, select any two distinct spherical tangent image plane pairs as the sensor directions, and exit; otherwise, go to Step 9.

Step 9—Reexamination: Find the optimal sensing direction and check if the enclosing sphere contains two other sensing directions such that every junction point of the monitored moving object appears on at least two of these three corresponding image planes in each of the F sampling periods. If so, then determine N_s (according to Lemma 11) to be 3, select these three directions as the sensor directions, and exit; otherwise, determine N_s finally (according to Theorem 4) to be 4, and select any two distinct spherical tangent plane pairs as the sensor directions. Exit.

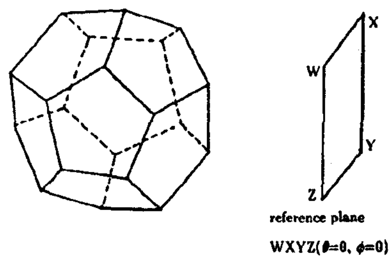
In Step 5, for every sampling period, the sensory system has to project the N junction points of the monitored object onto the $P \cdot Q$ quantized image planes. For the orthographic projection process including the hidden problem, the sensory system can use the hidden removing techniques provided in [27], [28] to determine the projective results on these $P \cdot Q$ quantized image planes over the F sampling periods.

V. Illustrative and Application Examples

A. Three Illustrative Examples

Three convex polyhedral objects imagined to move periodically in a 3-D environment are processed to verify the feasibility of the proposed algorithm. Here emphasis is put on the determination of the minimum number N_s of sensors.

Included in Fig. 5 is a simple convex polyhedral object repeating a sequence of line motions periodically, where the reference plane $WXYZ$ ($\theta = 0$, $\phi = 0$) is parallel to the object faces ABC and DEF, and perpendicular to all other faces of the object. The directions of the line motions are parallel to the object edge $\overline{CF}$. The repetitive motion procedure includes: 1) going to the left L meters; 2) going back to the original position; 3) going to the right L meters, and 4) finally going back to the original position again. By the proposed algorithm, it can be found that the enclosing sphere of the moving object contains many full sensing directions. The minimum number N_s of sensors is so determined to be two. Any two distinct ones of these full sensing directions, such as ($\theta = -\pi/4$, $\phi = 0$) and ($\theta = -3\pi/4$, $\phi = 0$), can be selected as the corresponding directions of the sensors.



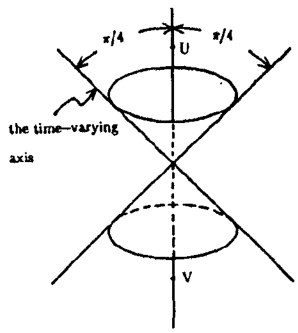


Fig. 7. Determination of minimum number N_s of sensors for moving dodecahedral object in 3-D space. (a) Moving dodecahedral object in 3-D space. (b) Time-varying axis in 3-D space.

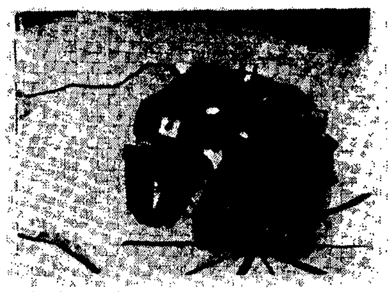


Fig. 8. Monitored robot with two sampling periods and eight feature points characterizing joint arm.

Illustrated in Fig. 6 is a cube repeating a sequence of line motions periodically, where the reference plane $WXYZ$ ($\theta = 0$, $\phi = 0$) is parallel to the object faces $ABCD$ and $EFGH$, and perpendicular to all other faces of the cube. The directions of the line motions are parallel to the object edge $\overline{DH}$ and its motion procedure is the same as that of the object shown in Fig. 5. By the proposed algorithm the enclosing sphere of the moving object contains no full sensing direction. However, there exist multiple optimal sensing directions. The total number of the junction points of the moving cube over all the F sampling periods visible from any optimal sensing direction (image plane) is equal to $7 \cdot F \cdot N/8$ which is larger than $2 \cdot F \cdot N/3$ (here, $N = 8$). By Step 8 of the proposed algorithm, N_s is determined to be either 3 or 4. By Step 9 of the proposed algorithm, an optimal sensing direction, say ($\theta = -\pi/4$, $\phi = \pi/4$), seeing the junction points A, B, D, E, F, G, H in every sampling period is first selected. Next, two other optimal sensing directions ($\theta = 3\pi/4$, $\phi = \pi/4$) seeing the junction points A, B, C, D, E, F, G in every sampling

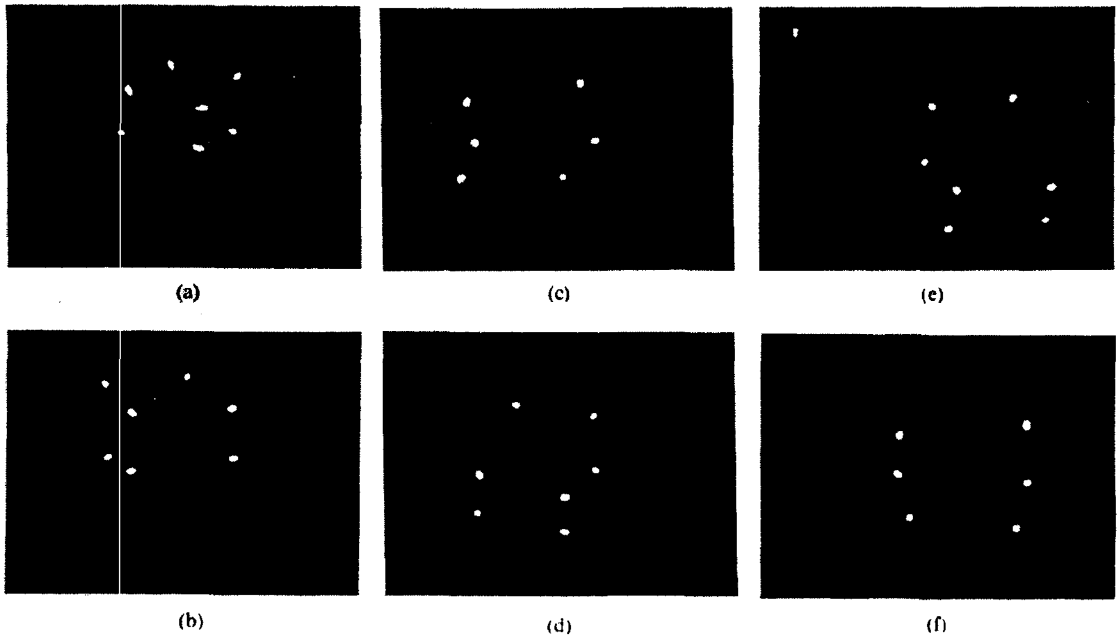


Fig. 9. Monitoring system with three cameras, $F=2$, and $N=8$. (a) Image of camera 1 at sampling time instant 1. (b) Image of camera 2 at sampling time instant 1. (c) Image of camera 3 at sampling time instant 1. (d) Image of camera 1 at sampling time instant 2. (e) Image of camera 2 at sampling time instant 2. (f) Image of camera 3 at sampling time instant 2.

period and ($\theta=-\pi/4$, $\phi=-\pi/4$) seeing the junction points A, C, D, E, F, G, H in every sampling period are selected. Since every junction point of the moving cube appears on at least two of these three selected image planes in every sampling period, the minimum number N_s of sensors is finally determined to be 3.

Fig. 7(a) includes a dodecahedral object repeating a sequence of rotational motions periodically. The object rotates periodically about a time-varying axis in a 3-D space as illustrated in Fig. 7(b), where the time-varying axis rotates about the fixed line $\overleftrightarrow{UV}$ with a slant angle $\pi/4$ and the line $\overleftrightarrow{UV}$ is parallel to the reference plane $WXYZ$ ($\theta=0$, $\phi=0$). By the proposed algorithm it is found that the enclosing sphere of the moving object contains no full sensing direction. The total number of junction points over the F sampling periods visible from the optimal sensing direction is also larger than $2\cdot F\cdot N/3$ (here, $N=20$). Thus N_s is determined to be either 3 or 4. By Step 9 of the proposed algorithm, an optimal sensing direction, such as ($\theta=-\pi/2$, $\phi=0$), is first chosen. However, it is impossible to find two other sensing directions such that every object junction point appears on at least two of the three selected image planes over all the F sampling periods. Other combinations of three image planes lead to similar results. Thus the minimum number N_s of sensors is finally determined to be 4. By Step 9 of the algorithm, any two distinct spherical tangent image plane pairs can be selected as the directions of the four sensors.

B. An Application Example

Here the images of a periodically working robot are processed to demonstrate the applicability of the set of proposed sensing strategies to monitoring an operating robot. The monitored robot is sampled with two sampling periods during a complete operation session. The monitored part is a joint arm of the monitored robot, as illustrated in Fig. 8. The joint arm is a convex polyhedral object, which can be characterized by eight feature points (i.e., $F=2$ and $N=8$). For experimental and demonstration convenience, light emitting diodes (LED's) attached to the junction points of the joint arm are used as the feature points for monitoring. Fig. 9 illustrates the six images of the three monitoring cameras over the two sampling periods, where Fig. 9(a)–(c) are the three images of the monitoring cameras of the first sampling time instant and Fig. 9(d)–(f) are the three images of the monitoring cameras of the second sampling time instant.

By Steps 7 and 8 of the proposed algorithm, it is found that there exists no full sensing direction and the total number (=14) of the junction points of the monitored polyhedral object over the two sampling periods visible from the optimal sensing direction is larger than $2\cdot F\cdot N/3$ (=32/3). Then, by Step 8 of the algorithm, the minimum number N_s of sensors necessary for monitoring the robot is determined to be either 3 or 4. Thus by Step 9 of the algorithm, N_s is determined finally to be 3.

VI. Conclusion

In this correspondence a set of new sensing strategies for monitoring 3-D moving convex polyhedral objects by computer vision is developed. The 3-D points are selected as the features for monitoring. It is proved that the minimum measurable feature point set for monitoring a 3-D moving convex polyhedral object is just the set containing all the junction points of the polyhedral object, and that the minimum data acquisition rate or the Nyquist rate for monitoring the object can be determined with discretely sampled 2-D image sequence data only. Several properties of orthographic projection for deciding the search space for the minimum number N_s of sensors necessary for monitoring the object are investigated. The bounds on N_s are also derived, and an algorithm for the determination of N_s and the corresponding directions of the sensors is proposed. The feasibility of the proposed algorithm is finally demonstrated by three illustrative examples and an application example.

In many applications it is necessary to monitor a more general object, such as an articulated robot arm which can be modeled as a joint-type curved object (i.e., an object containing many curved rigid bodies linked by joints). In this case, the convexity property on the object surface is lost, resulting in invisibility of certain object surface points from all directions at certain sampling time instants. Furthermore, it is also hard to find a sparse measurable feature point set for monitoring completely a general object in general motion conditions. Therefore, to monitor a general moving object is a different problem which should be treated from other viewpoints and is worth further research.

References

[1] L. A. Schmitt, W. A. Gruver, and A. Ansari, "A robot vision system based on two-dimensional object-oriented models," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-16, no. 4, pp. 582–589, 1986.

[2] W. F. Clocksin *et al.*, "An implementation of model-based visual feedback for robot arc welding of thin sheet steel," *Int. J. Robotics Res.*, vol. 4, no. 1, pp. 13–26, 1985.

[3] A. P. Ambler *et al.*, "A versatile computer-controlled assembly system," in *Proc. 3rd Int. Joint Conf. Artificial Intell.*, Stanford, CA, Aug. 1973.

[4] W. A. Perkins, "A model-based vision system for industrial parts," *IEEE Trans. Comput.*, vol. C-27, no. 2, pp. 126–143, 1978.

[5] U. L. Haass, "A visual surveillance system for tracking of moving objects in industrial workroom environments," in *Proc. 6th Int. Conf. Pattern Recognition*, Munich, Germany, 1982, pp. 757–759.

[6] G. G. Dodd and L. Rossol, *Computer Vision and Sensor-Based Robots*. New York: Plenum, 1979.

[7] I. Aleksander, *Artificial Vision for Robots*. New York: Chapman and Hall, 1984.

[8] K. C. Chang and W. H. Tsai, "3-D object inspection from multiple camera views," *Int. J. Pattern Recognition, Artificial Intell.*, vol. 1, no. 1, pp. 85–102, 1987.

[9] J. Y. S. Luh and J. A. Klaasen, "A three-dimensional vision by off-shelf system with multi-cameras," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 35–45, 1985.

[10] T. Hasegawa, "An interactive system for modeling and monitoring a manipulation environment," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-12, no. 3, pp. 250–258, 1982.

[11] M. O. Shneier, R. Lumia, and E. W. Kent, "Model-based strategies for high-level robot vision," *CVGIP*, vol. 33, pp. 293–306, 1986.

[12] H. Freeman, "Shape description via the use of critical points," *Pattern Recognition*, vol. 10, pp. 159–166, 1978.

[13] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, vols. I and II, 2nd ed. New York: Academic, 1982.

[14] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[15] W. E. L. Grimson, "Sensing strategies for disambiguating among multiple objects in known poses," *IEEE J. Robotics Automat.*, vol. RA-2, no. 4, pp. 196–213, 1986.

[16] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[17] A. Papoulis, *Systems and Transforms with Applications in Optics*. New York: McGraw-Hill, 1968.

[18] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[19] I. Carlbom and J. Paciorek, "Planar geometric projections and viewing transformations," *ACM Comput. Surveys*, vol. 10, no. 4, pp. 465–502, 1978.

[20] J. K. Aggarwal, L. S. Davis, and W. N. Martin, "Correspondence processes in dynamic scene analysis," *Proc. IEEE*, vol. 69, no. 5, pp. 562–572, 1981.

[21] E. L. Hall, D. L. Davies, and M. E. Casey, "The selection of critical subsets for signal, image, and scene matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 4, pp. 313–322, 1980.

[22] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. New York: Springer-Verlag, 1985.

[23] F. P. Preparata and S. J. Hong, "Convex hulls of finite sets of points in two and three dimensions," *Commun. ACM*, vol. 20, no. 2, pp. 87–93, 1977.

[24] S. Ganapathy, "Decomposition of transformation matrices for robot vision," in *Proc. Int. Conf. Robotics*, Atlanta, GA, 1984, pp. 130–139.

[25] M. P. Groover *et al.*, *Industrial Robotics: Technology, Programming, and Applications*. New York: McGraw-Hill, 1986.

[26] S. M. Udupa, "Collision detection and avoidance in computer controlled manipulators," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, 1977.

[27] I. E. Sutherland, R. F. Sproull, and R. A. Schumacker, "A characterization of ten hidden-surface algorithms," *ACM Comput. Surveys*, vol. 6, no. 1, pp. 1–55, 1974.

[28] J. D. Foley and A. Van Dam, *Fundamentals of Interactive Computer Graphics*. Reading, MA: Addison-Wesley, 1982.

[29] W. T. Miller, III, "Sensor-based control of robotic manipulators using a general learning algorithm," *IEEE J. Robotics Automat.*, vol. RA-3, no. 2, pp. 157–165, 1987.

[30] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill, 1977.

[31] D. F. Rogers and J. A. Adams, *Mathematical Elements for Computer Graphics*. New York: McGraw-Hill, 1976.

[32] D. F. Riddle, *Analytic Geometry With Vectors*, 2nd ed. CA: Wadsworth Publishing, 1977.

[33] P. S. Yeh, S. Barash, and E. Wysocki, "A vision system for safe robot operation," in *Proc. IEEE Int. Conf. Robotics and Automation*, Philadelphia, PA, 1988, pp. 1461–1465.

[34] K. Ikeuchi, "Recognition of 3-D objects using the extended Gaussian image," in *Proc. 7th Int. Conf. Artificial Intell.*, Vancouver, BC, Canada, 1981, pp. 595–600.

Segmentation Between Overlapping Parts: The Moving Shadows Approach

DANIEL RAVIV, MEMBER, IEEE, YOH HAN PAO, FELLOW, IEEE, AND KENNETH A. LOPARO, MEMBER, IEEE

Abstract —**A new method for segmenting three-dimensional overlapping surfaces is presented. It is based on moving a light source in a horizontal plane relative to the surfaces to be segmented. Using a camera which is placed above the surfaces, the shadows cast by the surfaces at each light source angle are recorded and analyzed. The segmentation algorithm is simple and based on Boolean processing of the data. A set of experimental results demonstrates the robustness and usefulness of the method.**

I. Introduction

One of the remaining problems in the robotics vision area is obtaining segmentation between overlapping surfaces. Many systems that have been successful in recognizing discrete parts have limitations making the recognition or segmentation of overlapping parts difficult. Many of the methods deal with two-dimensional (2-D) overlapping objects, attempting to find the overlapping surfaces; only few deal with three-dimensional (3-D) objects.

Ballard [1] developed a restricted form of the generalized Hough transform that can be used for recognizing partially hidden objects. Bolles and Cain [2] developed an approach for hidden object recognition referred to as "local feature focus." A technique for recognizing an object from a partially occluded boundary is given in [3]. A method for recognition of two overlapping parts using a single camera is introduced in [4]. A recent approach for recognition and positioning of a two-dimensional object is presented in [5]. A hand–eye system has been developed to perform bin picking [6]: photometric stereo vision is used to determine surface orientation. An experimental robot to acquire a class of workpieces from a bin using vision was demonstrated by Birk and Kelley [7]: a system was developed to pick up cylindrical parts. Other methods for segmenting images are described in [11]. However, most of them use gray-level images, emphasizing gray level and gray-level differences as indicators of segments. None of the known methods uses multiple binary images to obtain the segmentation.

In this correspondence, we describe a method for the segmentation of 3-D surfaces using one camera (see [8], [9]). The light source is rotated in a horizontal plane, and the camera is placed above the surfaces to be segmented. Points on the 3-D surfaces

Manuscript received April 15, 1988; revised July 29, 1988, and December 22, 1988. This work was supported in part by the Center for Automation and Intelligent Systems Research at Case Western Reserve University and in part by DARPA under Contract N00014-84K-0098. This material was partially presented at the SPIE Symposium on Advances in Intelligent Robotic Systems, Cambridge, MA, 1987.

D. Raviv is with the Department of Electrical Engineering, Florida Atlantic University, 500 N.W. 20th Street, Room 231, Boca Raton, FL 33431.

Y. H. Pao is with the Department of Electrical Engineering and Applied Physics, Glennan Building, Case Western Reserve University, Cleveland, OH 44106.

K. Loparo is with the Department of Systems Engineering, Case Western Reserve University, Cleveland, OH 44106.

IEEE Log Number 8926811.

0018-9472/89/0700-0880$01.00