OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Task-oriented generation of visual sensing strategies in assembly tasks
Author(s)	三浦, 純; Ikeuchi, Katsushi
Citation	IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998, 20(2), p. 126-138
Version Type	VoR
URL	https://hdl.handle.net/11094/3099
rights	©1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE..
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Task-Oriented Generation of Visual Sensing Strategies in Assembly Tasks

Jun Miura, *Member, IEEE*, and Katsushi Ikeuchi, *Fellow, IEEE*

Abstract—This paper describes a method of systematically generating visual sensing strategies based on knowledge of the assembly task to be performed. Since visual sensing is usually performed with limited resources, visual sensing strategies should be planned so that only necessary information is obtained efficiently. The generation of the appropriate visual sensing strategy entails knowing what information to extract, where to get it, and how to get it. This is facilitated by the knowledge of the task, which describes what objects are involved in the operation, and how they are assembled. In the proposed method, using the task analysis based on face contact relations between objects, necessary information for the current operation is first extracted. Then, visual features to be observed are determined using the knowledge of the sensor, which describes the relationship between a visual feature and information to be obtained. Finally, feasible visual sensing strategies are evaluated based on the predicted success probability, and the best strategy is selected. Our method has been implemented using a laser range finder as the sensor. Experimental results show the feasibility of the method, and point out the importance of task-oriented evaluation of visual sensing strategies.

Index Terms—Task-oriented vision, sensing planning, active vision, CAD-based vision, vision-based assembly.

1 Introduction

In vision-guided robotic operations, vision is used for extracting necessary information for proper task execution. Since visual sensing is usually performed with limited resources, visual sensing strategies should be planned so that only necessary information is obtained efficiently. To determine an efficient visual sensing strategy, knowledge of the task is necessary. Without knowledge of the task, it is often difficult to select the appropriate visual features to be observed. In addition, resources may be wasted in tracking uninformative features.

From this standpoint, research on *task-oriented vision*, *active vision*, or *purposive vision* has been actively investigated [1], [2], [4], [5], [13]. By using knowledge of the task, the vision system can be designed to be fast and robust. However, the designing process tends to be task-specific and requires a significant amount of effort. Thus, it is desirable to develop a systematic method which can generate *task-oriented* visual sensing strategies automatically, namely a method that optimizes each visual sensing strategy according to a given task.

The generation of task-oriented sensing strategy is decomposed into the following three subproblems to be solved successively:

- determine what visual information is necessary for the current task;
- determine which visual features carry such necessary information; and
- determine how to get necessary information with the sensors used.

The first two subproblems are concerned with focusing the attention to informative visual features; the last problem is concerned with evaluation of sensing strategies.

The ability of focusing attention is important to realize efficient visual sensing strategies [28]. There have been several approaches to this problem.

Hutchinson and Kak [12] dealt with the problem of resolving the ambiguity of sensor information. They used Dempster-Shafer theory to represent uncertainties of hypotheses in object identification. An entropy of a set of hypotheses was used as a utility function; a sensor placement was selected which minimizes the entropy.

Rimey [22] presented a framework of task-oriented vision which can solve high-level vision problems such as determining which object to search for next to answer a query. The knowledge of the task is represented by a Bayesian network, and the sensing action is selected which has the highest expected utility. The utility function is defined as the combination of the predicted information value and the sensing cost.

Birnbaum et al. [6] presented a vision system which can explain a scene of blocks world in terms of stability of block structures. Using the rules derived from causal knowledge of naive physics, the focus of attention is moved to look for evidence that explains the situation.

These works are concerned with exploratory visual sensing tasks under uncertainty of the knowledge of the scene. The visual feature set, from which the observed features are selected, is given in advance; it is not automatically derived from the task description.

- *J. Miura is with the Department of Computer-Controlled Mechanical Systems, Osaka University, Suita, Osaka 565, Japan. E-mail: jun@mech.eng.osaka-u.ac.jp.*
- *K. Ikeuchi is with the Institute of Industrial Science, The University of Tokyo, 7-22-1, Roppongi, Minato, Tokyo 106, Japan. E-mail: ki@cvl.iis.u-tokyo.ac.jp.*

Manuscript received 17 Feb. 1997; revised 4 Dec. 1997. Recommended for acceptance by P. Flynn.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106039.

0162-8828/98/$10.00 © 1998 IEEE

Kuniyoshi and Inoue [17] proposed a framework of qualitatively recognizing ongoing human action. Using a hierarchical action model, which is given in advance, possible upcoming events are predicted, and visual features to observe are selected based on that prediction.

Horswill [10] proposed a concept of *specialization* for constructing task-specific robot vision systems. By analyzing the property of the task, including the environment in which the robot operates, a simple but robust vision algorithm is organized from a set of given small vision processes.

These works are concerned with the usage of the task, including the constraints on the environment, for concentrating the visual processing to only the necessary portion of image. This allows visual recognition to be fast and robust. These approaches are, however, highly task-specific and are based on the careful a priori examination of the task.

For sensor planning in inspection tasks, several methods have been proposed which generate a set of features to be observed. Features are indicated directly in the inspection specification [25] or are selected from the specification of entities to be measured through given knowledge of the mapping from measurable entities of an object to features to be observed [30]. In sensor planning for inspection, derivation of the feature set to be observed is relatively easy, because the purpose of the task itself is visual recognition.

Automatic generation of recognition programs can be viewed as sensing strategy planning. Ikeuchi has been developing Vision Algorithm Compiler (VAC) [8], [14], which can generate object recognition programs using explicit models of both objects and sensors. A VAC analyzes the appearances of objects using the models, and generates object recognition programs, usually in the form of decision trees.

The third subproblem (i.e., how to get necessary information) is decomposed further into two more specific problems of determining a set of feasible sensing strategies and subsequently selecting the most appropriate one among them. The goal of the former is to determine sensing condition which satisfies several requirements on imaging such as resolution, field of view, focus, and visibility [7], [23], [26], [27]. The ability to solve such a problem would be necessary for any sensor planners as a subroutine of automatically determining feasible sensing conditions.

As mentioned above, the second more specific problem is to determine the best sensing strategy which maximizes some "goodness" function. The minimum uncertainty criterion has often been used [16], [29], [31]; some measure of uncertainty, such as the determinant of the covariance matrix of the parameter vector to be estimated, is used for selecting the best strategy. In certain types of tasks, however, this criterion may not be appropriate; some part of information may need to be more accurate than the rest for a specific task, for example. A weighted sum of uncertainty parameters is one way to deal with such a case. It is, however, difficult to determine appropriate weights for a given task. Thus, some appropriate function should be automatically designed for each task which can measure how each sensing strategy contributes to the proper execution of the current task.

This paper proposes a novel method of systematically generating visual sensing strategies based on knowledge of the task to be performed. We deal with visual sensing strategy generation in assembly tasks, in which the environment is known, that is, the shape, the size, and the *approximate* location of every object is known to the system. In this situation, the role of visual sensors is to determine the position of the currently assembled object with sufficient accuracy so that the object can be, with a high degree of certainty, mated with other objects. The proposed method generates optimal sensing strategies by solving the three subproblems mentioned above.

We have been developing a novel robot programming system, the APO (Assembly Plan from Observation) system [15]. The system generates the description of an assembly task by observing human performance of the task. The task description is then mapped into an actual robot to perform the same task. Although the APO system cannot provide direct sensing strategy from observing human actions, useful information for generating visual sensing strategies is automatically obtained from analysis of the observed tasks.

In assembly operations, DOFs of assembled objects are gradually constrained. Thus, specific DOFs (*critical dimension*) of the objects need to be observed in each assembly operation. Section 2 briefly explains the face contact analysis of assembly operations in the APO system [15] and then describes the method of deciding, based on the face contact analysis, whether the visual feedback is necessary for the current assembly operation. This analysis also provides the critical dimensions to be observed (*what information to extract*). Section 3 presents a method to determine necessary features to be observed through the analysis of CAD models and the critical dimensions (*which features to observe*). Section 4 provides the optimal sensing strategy to observe such features (*how to observe the features*), and Section 5 describes an implementation of the sensing strategies and their performance evaluations using a line laser range finder. Section 6 summarizes the paper and discusses an extension of our approach to more general visual sensing strategy generation.

2 Determining What Information Is Necessary

This paper assumes that a robot has the capability to perform passive compliant motions [19], which are the motions to keep the current physical contacts between objects using force information. Under this assumption, this section examines which assembly operations really require visual feedback, and which operations can be performed only with passive compliant motions. The analysis is based on face contact relation transitions between the manipulated and the environmental objects. First, we describe the analysis results for polyhedral objects with three translational motions [15]; then, we extend the analysis to include cylindrical objects with four dimensional motions: three translational motions and one rotational motion.

2.1 Face Contact Relations

We assume that each assembly operation involves one manipulated object and several stationary environmental objects

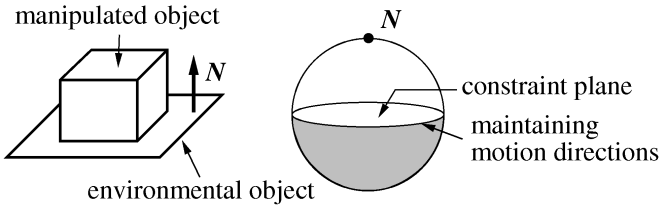


Fig. 1. Constraint inequality depicted on the Gaussian sphere.

that have face contacts with the manipulated object. Face contact relations are defined between the manipulated and the environmental objects.

Let us suppose a face of the manipulated object has a face contact to a face of an environmental object (see Fig. 1). Each face contact pair constrains the possible translation motions of the manipulated object: $\boldsymbol{N} \cdot \Delta \boldsymbol{T} \geq 0$, where $\Delta \boldsymbol{T}$ denotes possible translational vectors of the manipulated object, and $\boldsymbol{N}$ denotes the normal of the contact face.

We use points on the Gaussian sphere to specify both a constraint vector and all possible motion vectors. Each motion unit vector is translated so that its start point is located at the center of the Gaussian sphere and its end point exists at some point on the surface of the Gaussian sphere. We use this end point to denote the vector.

The constraint given by a contact face pair defines several regions in the Gaussian sphere. We refer to the plane perpendicular to the normal, $\boldsymbol{N}$, as the constraint plane; this plane divides the Gaussian sphere into two hemispheres. Without loss of generality, assuming that the normal points to the north pole of the Gaussian sphere, the northern hemisphere corresponds to possible translational motion directions; the southern hemisphere corresponds to prohibited motion directions toward which the manipulated object cannot move.

In Fig. 1, motion directions corresponding to the boundary between the hemispheres (the equator) maintain the current face contact between the manipulated and environmental objects. Those directions and those DOFs are referred to as the maintaining directions and the maintaining DOF, respectively. In this example, maintaining DOFs are two. Motions of the directions corresponding to the inside of the detaching hemisphere break the face contact, and are referred to as detaching motions. In this example, a pure detaching motion, which does not contain any maintaining motion component, is toward the north pole; it has one DOF (detaching DOF).

When several surface patches of different orientations make contact, possible motion directions are constrained through simultaneous linear inequalities. These constraints are represented as a combined region in the Gaussian sphere.

Fig. 2 shows the case where two normal vectors of environmental objects have the opposite directions. The possible motion directions of the manipulated object can be represented as the entire great circle perpendicular to the axis connecting the two poles. These motions have two DOFs and are maintaining motions. There are no detaching motions; the detaching DOF is zero. The direction along the surface normal is completely constrained; the DOFs of the constraint directions (constraining DOF) is one. Note that the sum of the maintaining DOF, the detaching DOF, and the constraining DOF is three, the entire DOF of three-dimensional translational motions.

We can specify face contact relations by using a triplet of maintaining, detaching, and constraining DOFs. For example, using this triplet, the relations of Fig. 1 and Fig. 2 are represented as (2, 1, 0) and (2, 0, 1), respectively.

In general, possible contact relations are classified into 10 contact relations as shown in Fig. 3, [15]. The triplet of DOFs for each relation is also indicated in the figure.

Each assembly step causes a *transition* from one face contact relation to another. Extracting the admissible transitions results in the contact state transition graph as shown in Fig. 4. Note that it was early assumed that there was only *one* manipulated object and several stationary objects, and that there can thus be no transition to or from state I:(0 0 3).

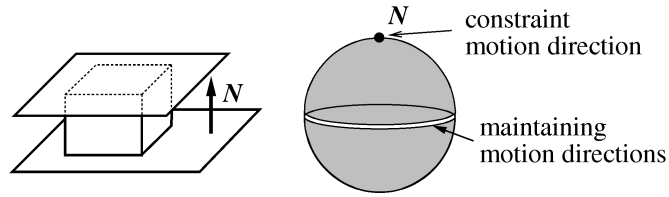


Fig. 2. A bidirectional constraint.

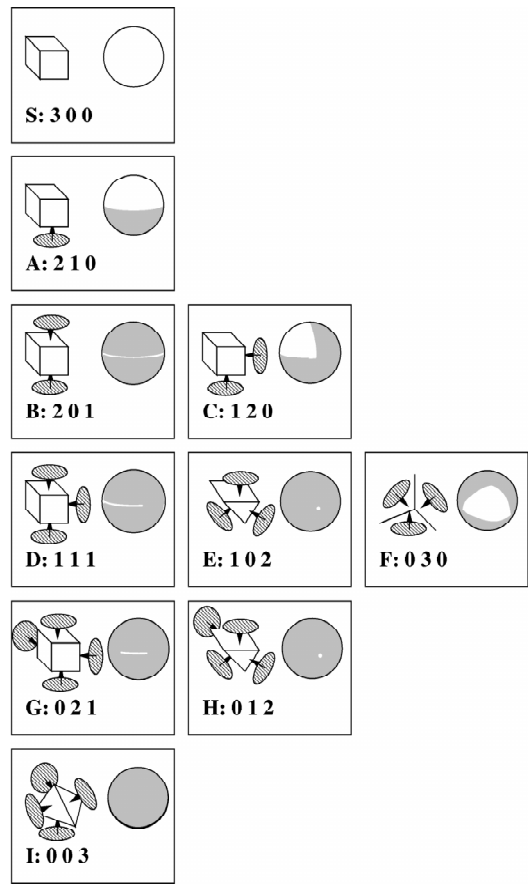


Fig. 3. Ten contact states [15]. The white areas in the Gaussian sphere denote possible motion vectors. Each state has a label. The three digits denote maintaining DOF, detaching DOF, and constraining DOF, respectively.

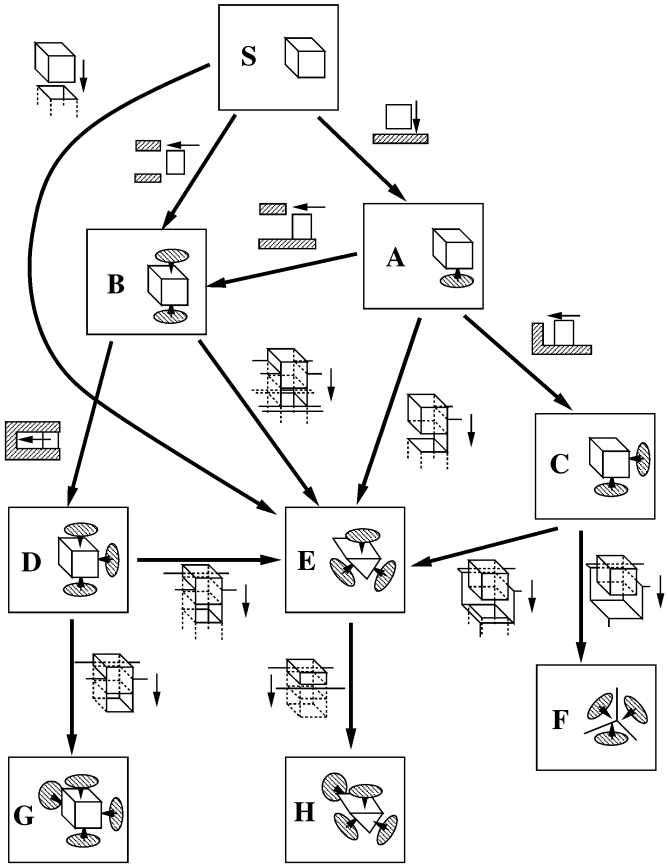


Fig. 4. Contact state transitions represented as a directional graph [15].

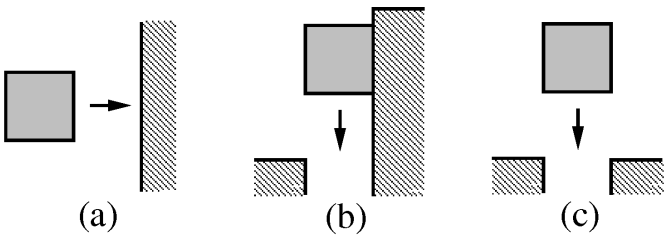


Fig. 5. Three typical cases of increase of constraint on a degree of freedom. Type and transition of the triplet is as follows: (a): maintaining → detaching ((3,0,0) → (2,1,0)). (b): detaching → constraining ((2,1,0) → (2,0,1)). (c): maintaining → constraining ((3,0,0) → (2,0,1)).

2.2 DOF Transitions and Necessary Visual Information

An assembly operation (i.e., a transition of a face contact state) always increases constraints on some DOFs of the manipulated object. This increment is classified into three classes: from maintaining DOF to detaching DOF; from detaching DOF to constraining DOF; and from maintaining DOF to constraining DOF. Fig. 5 shows typical examples of the three classes.

Let us examine the type of the DOF transition in examples in Fig. 5. In Fig. 5a, the horizontal operation translates the maintaining DOF (horizontal freedom) into the detaching DOF. The approaching direction of the block is parallel to the direction of the pure detaching motion at the goal relation A, i.e., the normal vector against the wall. This class of operations can be performed by a compliant motion without visual feedback, such as a move-to-contact operation.

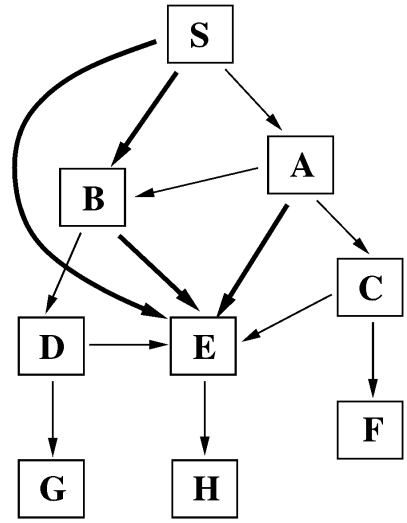


Fig. 6. Classification of state transitions. Bold lines indicate the transitions that require visual information. Thin lines indicate the transitions that do not require visual information.

In Fig. 5b, the vertical operation translates the detaching DOF of the horizontal direction to the constraining DOF. A compliant motion maintains the contact between the block and the right wall and achieves the desired horizontal position. The final constraining DOF is automatically achieved by this compliant motion, e.g., move-with-maintaining-contact, without visual feedback.

In Fig. 5c, the vertical operation translates the maintaining DOF of the horizontal direction to the constraining DOF. The horizontal position of the block needs to be adjusted before mating, so that both the left and the right face contacts are achieved simultaneously. Before this operation, along the horizontal maintaining degree, a manipulated object has no physical contact; after this operation, the constraining degree from both walls occurs simultaneously. Thus, simple compliant motion cannot achieve such translation. This operation needs visual feedback; this degree, which is referred to as *critical dimension*, should be observed by visual feedback.

We can summarize the above arguments into the following criteria:

1) maintaining DOF to detaching DOF: no visual feedback is necessary.
2) detaching DOF to constraining DOF: no visual feedback is necessary.
3) maintaining DOF to constraining DOF: visual feedback is necessary.

Applying these criteria to 13 admissible transitions in Fig. 4 provides four transitions that require visual feedback (S-to-B, S-to-E, A-to-E, and B-to-E) indicated with bold lines in Fig. 6.

2.3 Extension of Face Contact State

So far, the analysis covers only three-dimensional translation motions with planar face contacts. We will extend the analysis to include one additional rotational motion about

the z axis. This extended analysis denotes face contact relations with a sextuplet of DOFs: one triplet for translational DOFs and another triplet for rotational DOFs.

We also include cylindrical objects for the analysis. We define the three classes of contact relations between cylindrical surfaces as shown in Fig. 7. No contact, half contact, and full contact in Fig. 7 correspond, respectively, to maintaining DOF, detaching DOF, and constraining DOF with respect to the translational motion perpendicular to the axis of rotation. Since this paper deals only with face contact relations, we do not consider cylindrical-planar contacts.

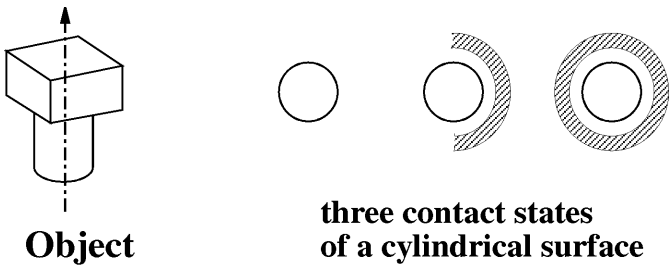


Fig. 7. The object considered in the contact state analysis.

This extension allows the system to handle a large class of realistic operations with industrial parts, including gears mating and bolt-nut operations, as shown by experiments later. It is also true that our Robotworld testbed has these four DOFs. Extension to point and/or line contact cases, or to general six DOF cases, will be discussed in Section 6.

Fig. 8 summarizes the analysis result. Here, the possible face contact relations are arranged according to their sextuplets of DOFs.

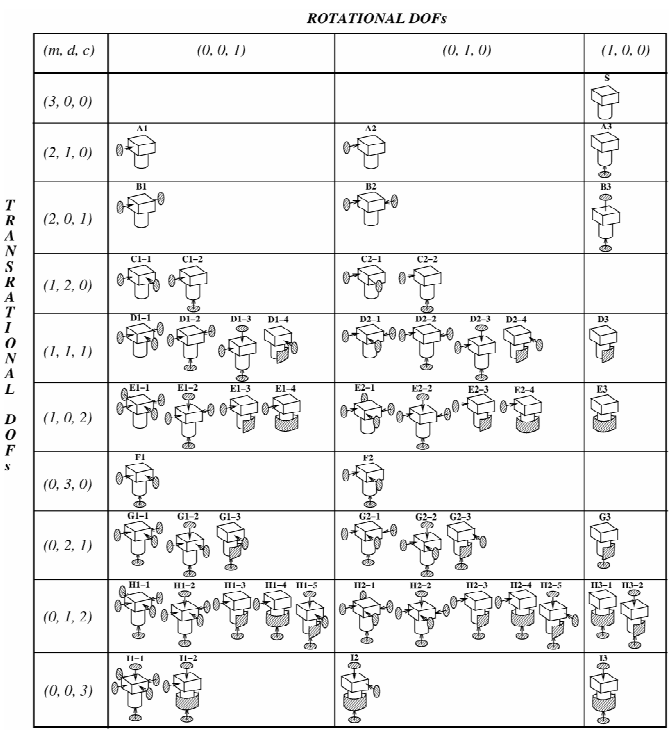


Fig. 8. Summary of extended contact state analysis.

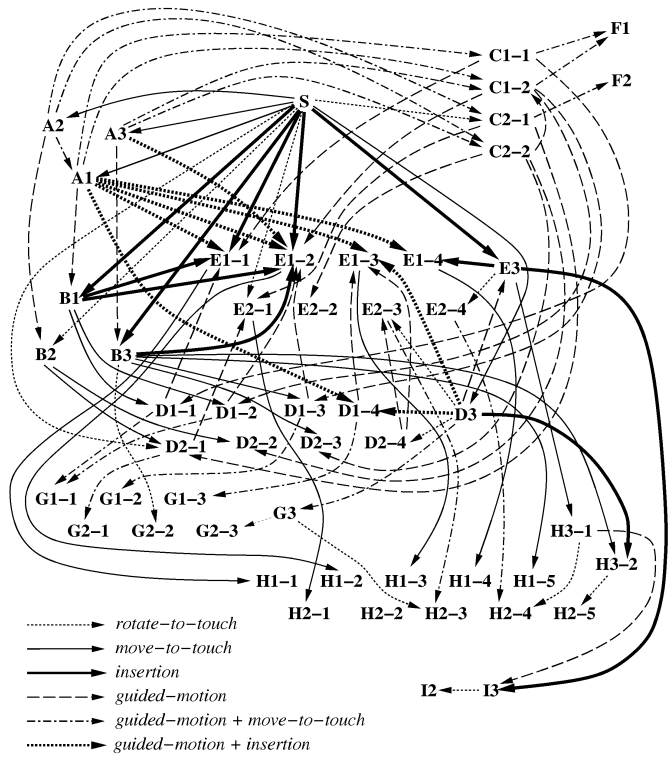


Fig. 9. Transition graph for extended analysis.

We enumerate possible transitions between contact relations. Eighty-five admissible transitions are extracted, as shown in Fig. 9, by applying the selection criteria for admissible transitions to all possible transitions. Labels in the figure (e.g., A1) are defined in Fig. 8.

If the number of allowable states increases, the transition graph is expected to be more complicated and to require more efforts to generate. Once the graph is generated, however, the complexity of the graph is not a serious problem, because the graph is used not for generating a sequence of assembly operations, but for detecting state transitions that require visual feedback, or for recognizing human operations by matching the recognition result with the graph [15].

We consider an infinitesimal rotation. Such an infinitesimal rotational motion is equivalent to an infinitesimal translational motion in terms of face contact relation transition. We can use the same criteria for selecting assembly operations that require visual feedback in this four-dimensional case.

By examining the sextuplet transitions in 85 cases, 19 transitions are selected as those that require visual feedback. These selected transitions are as shown with bold solid and bold broken arcs in Fig. 9. We can group these 19 cases into six groups with respect to the DOFs necessary to be observed. Fig. 10 shows the representative examples of these six groups.

3 Feature Selection

This section describes how to select a set of features that carries sufficient dimensional information for executing the

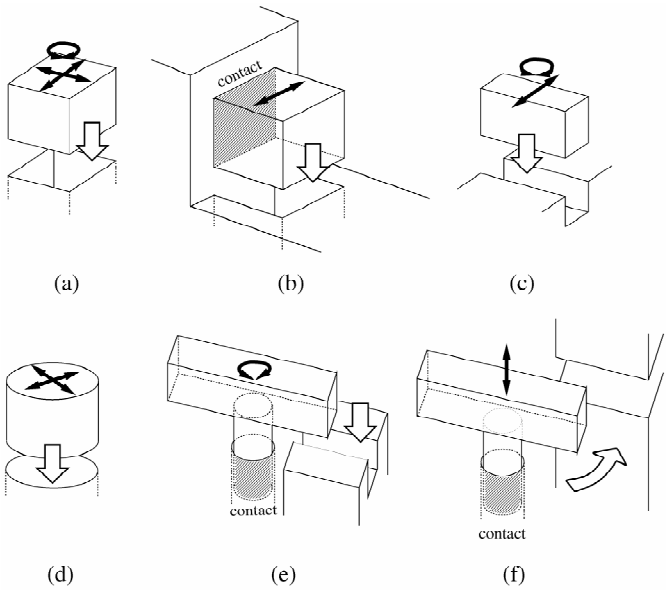


Fig. 10. Transition groups which need visual information. Thick arrows indicate the direction of movement. Thin arrows indicate degrees of freedom to be adjusted by use of visual information. The transition of sextuplet for each case is as follows: (a): (3,0,0;1,0,0) → (1,0,2;0,0,1). (b): (2,1,0;0,0,1) → (1,0,2;0,0,1). (c): (3,0,0;1,0,0) → (2,0,1;0,0,1). (d): (3,0,0;1,0,0) → (1,0,2;1,0,0). (e): (1,0,2;1,0,0) → (1,0,2;0,0,1). (f): (1,0,2;1,0,0) → (0,0,3;1,0,0).

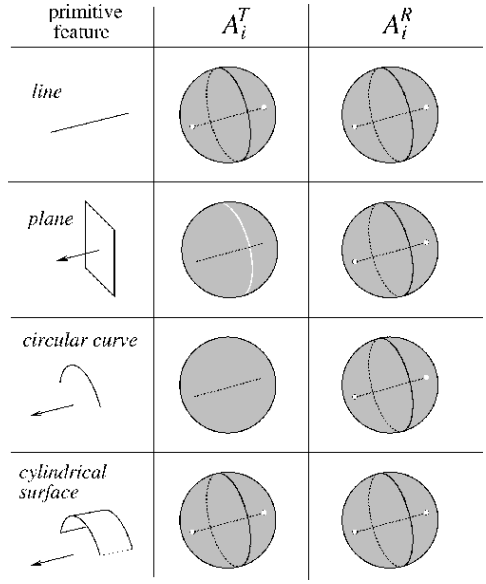


Fig. 11. Example sensing primitives represented by the Gaussian spheres.

operation with visual feedback resulting in the transition from maintaining DOF to constraining DOF.

3.1 Sensing Primitive

A sensing primitive describes the relationship between an observable feature, such as edges or planar faces, and its DOFs obtainable by observing it. Examples of sensing primitives are shown in Fig. 11. For example, by observing a straight line in 3D space, all the translation freedoms except those along the straight line are constrained. This constrained area, A_i^T, given by this primitive feature is denoted as the shaded area on the Gaussian sphere. All the rotations except that about the line are also constrained; the entire sphere except for a pair of points is constrained (see A_i^R in Fig. 11).

3.2 Feature Selection Process

The analysis of face contact relation transition by the current task provides constraining DOF added by this task. A set of observable features is obtained by examining the current CAD-based internal-world model. By consulting prepared sensing primitives with the observable features, a set of features required to be observed is obtained.

As an example of this process, let us consider case (a) peg-in-hole operation shown in Fig. 12. For the sake of simplicity, we will consider the case (a)'s translation operations. Before the insertion operation, the manipulated object has a full three-dimensional maintaining DOF. After this operation, the manipulated object has an equator of constraining DOF. Thus, the (a) insertion operation translates two-dimensional maintaining DOF perpendicular to the insertion direction to two-dimensional constraining DOF, referred to as critical DOF to be monitored with visual feedback. Fig. 12a depicts the full equator on the Gaussian sphere to depict these DOF.

Before case (b) operation, the manipulated object has a two-dimensional maintaining DOF as an equator perpendicular to the normal of the contact face. The constraining DOF after the operation is the same as in the (a) operation. Thus, the (b) operation translates one-dimensional maintaining DOF along the x axis to one-dimensional constraining DOF, which is indicated as a pair of points on the equator in Fig. 12b. The visual feedback is necessary along the x axis, but is not necessary along the y axis.

Fig. 13 shows two sensing primitives. By observing edge e_1, all the translation motions, except those along the y axis are constrained. Edge e_2 also provides similar constraints. In order to cover the full circle of the critical DOF in case (a) in Fig. 12, both e_1 and e_2 are necessary to be observed. On the other hand, only e_1 is sufficient for case (b) to cover the pair of points of the critical DOF.

This procedure can be formalized as follows: Since constraints for translational operations and that for rotational operations can be considered separately, we take two Gaussian spheres, G^T and G^R, and use points on each sphere to

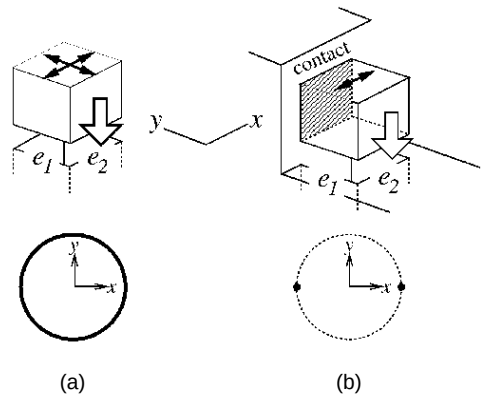


Fig. 12. Two insertion operations and dimensions to be monitored.

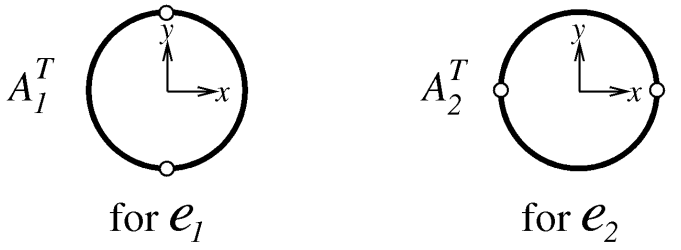


Fig. 13. Constraints obtained by observing one of the edges.

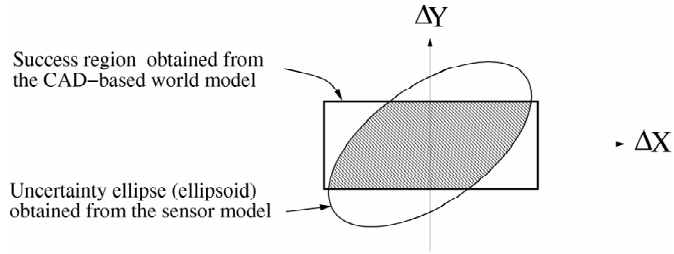


Fig. 14. Calculation of the predicted success probability. This figure shows the case where the position parameter of the object is two-dimensional, (X, Y).

represent constraints on translations (on G^T) or on rotations (on G^R). Let us denote:

- $A^T_{critical}\left(A^R_{critical}\right)$ as a set of points on G^T (G^R) to be translated from maintaining DOF to constraining DOF (critical DOF).
- $A_i^T\left(A_i^R\right)$ as a set of points on G^T (G^R) which represents constraints given by the observation of the ith sensing primitive.

By observing n different sensing primitives, the following two resultant point sets, $A^T_{constrained}$ and $A^R_{constrained}$, are obtained:

$$A^T_{constrained} = \bigcup_{i=1}^{n} A_i^T ,$$

$$A^R_{constrained} = \bigcup_{i=1}^{n} A_i^R .$$

In order that this set of features provides sufficient constraints, the resultant area should cover the critical DOF. Namely, the following condition must be satisfied:

$$A^T_{critical} \subseteq A^T_{constrained}$$

and

$$A^R_{critical} \subseteq A^R_{constrained} .$$

The equality signs indicate that exactly the required information is provided by observation, whereas $A_{critical} \subset A_{constrained}$ indicates there is more information than required.

4 Evaluation of Sensing Strategies

Once a set of features to be observed is selected, a set of possible sensor positions, from which all of the selected features are observable, is enumerated [7], [23], [26]. We determine the optimum sensing position among those possible sensing positions by evaluating them with respect to accuracy of the estimated object position. Evaluation methods based on covariance matrix are relatively common in unconstrained environments such as navigation application [3], [29]. Such covariance methods are not so satisfactory to reflect the relative importance of sensing dimensions given by the relationship between faces of manipulated and environmental objects. Thus, we decided to develop our novel method based on an operation's success probability.

4.1 Predicted Success Probability

A poor positional estimation of a manipulated object provides poor operational performance; an accurate estimation provides high performance. Thus, the predicted success probability of a current operation provides a measure of accuracy of the sensing strategy.

The success probability is determined by three steps. The first step calculates a success region, the subspace of the positional parameters in critical DOF such that if the uncertainty is inside the space, the current operation succeeds. The rectangular area in Fig. 14 is an example of the success region. The relative tolerance of the manipulated and environmental objects along the critical dimensions, given by the analysis of CAD models, provides the area. The second step calculates the uncertainty distribution of the measurements given by the current sensing strategy (the ellipse area in Fig. 14). The final step obtains the predicted success probability by the intersection between the success region and the uncertainty distribution of the measurements. This success probability is numerically calculated by quantizing the space of the positional parameters.

The predicted success probability is a general criterion for ranking sensing strategies, and is applicable to any sensors as long as the probabilistic model of uncertainty is provided; in addition, the uncertainty does not necessarily need to be modeled with Gaussian.

After calculating the success probabilities for all feasible sensing strategies, the one with the highest probability is selected as the optimal sensing strategy.

4.2 Success Region

A success region is formed in a space of critical DOF, those translated from maintaining to constraining DOF, in each operation. For example, the (a) peg-in-hole operation in Fig. 12 has three DOFs of critical DOF, two translational on the plane perpendicular to the insertion direction and one rotational around the insertion direction. A success region is a representation of the clearance of the operation in the critical DOF from the CAD models of objects; it can be calculated as a free area in the configuration space [18].

Since the current system consists of the assembly operations with planar or cylindrical surfaces, the following three cases are sufficient for consideration:

1) insertion of a polygonal cross-section peg with a hole (in Fig. 10a and Fig. 10b);
2) insertion of a peg with a polygonal cross-section into a parallel gap (in Fig. 10c, Fig. 10e, and Fig. 10f); and

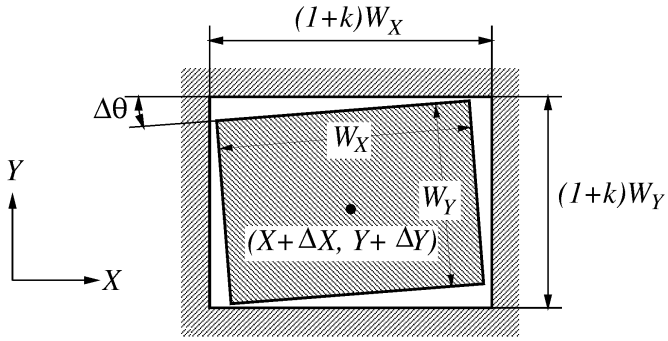
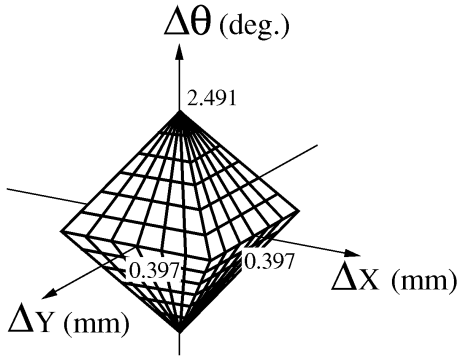
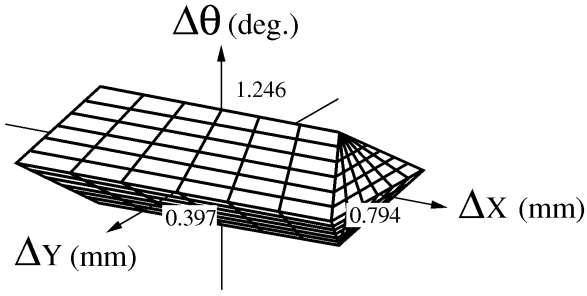


Fig. 15. Top view of rectangular peg-in-hole operation.



(a) Wx=19.05(mm), Wy=19.05(mm), k=0.043



(b) Wx=38.10(mm), Wy=19.05(mm), k=0.043

Fig. 16. Example success regions.

3) insertion of a peg with a circular cross-section into a hole (in Fig. 10d).

If there are some constraints before the operation (e.g., in Fig. 10b), the actual success region is given as a cross-section of the general success region cut by fixed parameter values under the constraints.

Here, as an example, we derive the success region for the peg-in-hole operation shown in Fig. 15. Edges of the hole are aligned to the X and the Y axes. Let W_X and W_Y be the widths of the peg in the X and Y axes, respectively. Also, let k denote the clearance ratio of the hole. These values come from the CAD model. We need to adjust the position and the orientation of the peg, (X, Y, θ).

We calculated the actual success regions for two sets of geometric values. Fig. 16 shows the resultant success regions.

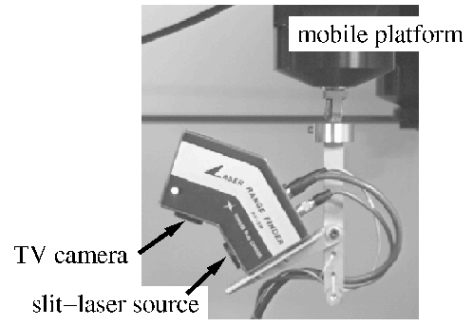


Fig. 17. Laser range finder.

As shown in the figure, the tolerance in ΔX in case (b) is larger than that in case (a), while the tolerance in $\Delta\theta$ is smaller. If the uncertainty distribution of the position parameter is the same in both cases, the resultant success probabilities should differ from each other because of different success regions. Thus, the effect of the uncertainty in the parameter vector to the task execution needs to be evaluated by considering the success region.

5 Experiments

5.1 Laser Range Finder

The proposed sensing method has been implemented using a Toyota line laser range finder (LRF) as the sensor (see Fig. 17) [20]. The LRF emits a slit laser, detects the highlighted portion of the object with a TV camera, and obtains a line of 3D measurements. The LRF is attached to one of the arms in the RobotWorld [24]. All the arms have four DOFs: three DOFs for translation and one DOF for rotation about the vertical axis (the z axis).

All assembly operations that require visual feedback belong to the "peg-in-hole" class operation. The location of a peg is measured by observing its side faces; the location of the hole is measured by observing several points (currently, five points) on its edges. Thus, we prepare sensing primitives for the following four geometric features: a straight edge, a circular edge, a planar face, and a cylindrical surface.

We use a general sensing strategy that measures by moving the LRF in parallel with the insertion direction (see Fig. 18); the relative displacement on the plane perpendicular to the insertion direction is important for the operation. We also control the position of the range finger so that each measured point is within a certain area of the slit laser; the uncertainty of the measurement with the LRF is considered to be constant in this area. Thus, the only parameter that specifies the position of the range finder is the angle between the direction of the laser and some axis of the plane perpendicular to the insertion direction (see Fig. 19).

The actual assembly operations with visual feedback are implemented to perform in a "stop and sense" mode. First, a peg is moved by a manipulator to the position just before a hole. Then, the LRF is placed in the planned position and measures the position of the hole and the peg. If the error in the relative position between the peg and the hole is within

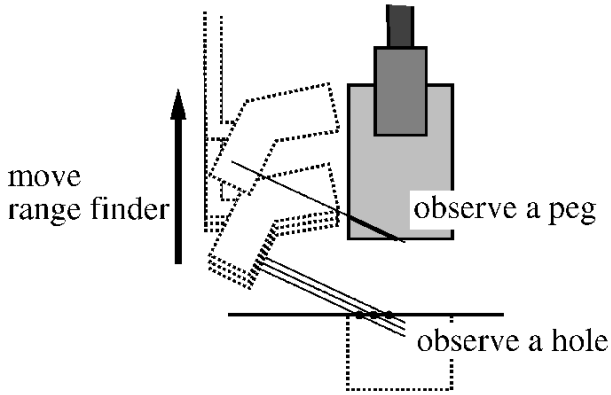


Fig. 18. A strategy for observing a peg and a hole.

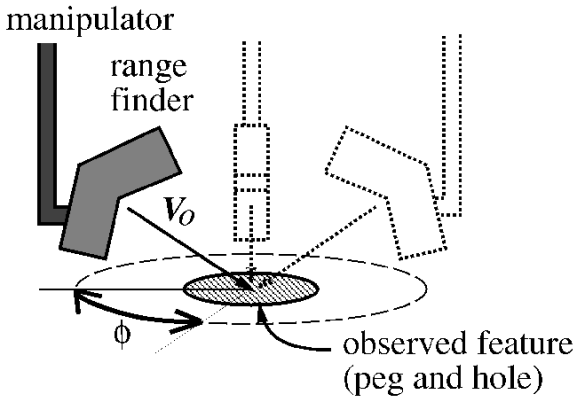


Fig. 19. Candidate positions.

the success region, the peg is inserted. Otherwise, the peg position is adjusted and the peg is observed again. This final step is repeated until the relative position becomes satisfactory, and then the peg is inserted.

For evaluation of the proposed method, we conducted three operations: a peg-in-hole operation, an operation that consisted of putting the tip of a screwdriver into the slot of a screw head, and a gear-mating operation. We then compared the success probability predicted by the object models and sensor models, with the actual success ratio, obtained through 50 trials of the operation by the robot arm and the sensor.

5.2 Uncertainty Model

Our laser range finder provides measurements with accuracy better than 0.1 [*mm*] in depth and better than 0.3 [*mm*] in the horizontal position. The depth and the horizontal position are measured on the laser plane. The purpose of this paper is, however, not to construct an uncertainty model of our laser range finder, but to demonstrate that our method can generate the optimal sensing strategy if the uncertainty model of the sensor is given. Thus, we artificially added a relatively large Gaussian noise to the measurement; we added a Gaussian noise with a standard deviation 0.12 [*mm*] to the depth measurement, and a Gaussian noise with a standard deviation 0.30 [*mm*] to the horizontal position measurement; these two Gaussian noises are set to be independent of each other.

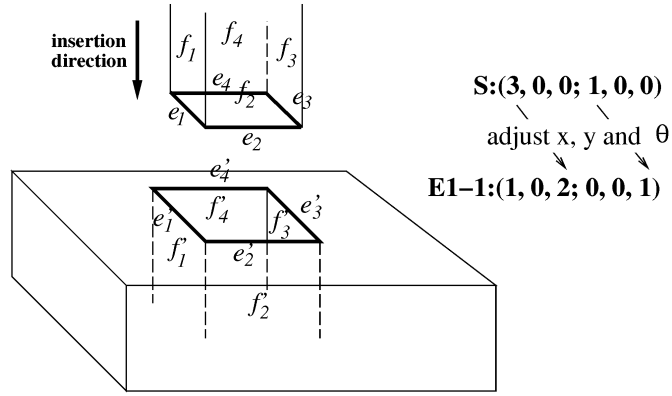


Fig. 20. Face contact analysis of the rectangular peg-in-hole operation. The sextuplet of DOFs (see Section 2 to Section 2.3) changes from (3,0,0; 1,0,0) to (1,0,2; 0,0,1).

5.3 A Peg-in-Hole Operation

5.3.1 Face Contact Analysis

The peg-in-hole operation inserts a rectangular-cross-sectional peg into a hole of the same shape. This operation belongs to group (a) in Fig. 10; the state transition is from (3,0,0; 1,0,0) to (1,0,2; 0,0,1). The critical DOF consists of two translational and one rotational degree. These dimensions will be monitored through visual feedback.

Let us consider the case in Fig. 20. This insertion operation establishes the following face contact: $(f_1 - f_1')$, $(f_2 - f_2')$, $(f_3 - f_3')$, and $(f_4 - f_4')$. The candidate features for observation include f_1, f_2, f_3, and f_4 for the peg, and e_1', e_2', e_3', and e_4' for the hole. The sensing primitive analysis indicates that observing two neighboring faces, such as f_1 and f_2, and edges, such as e_1' and e_2', provides sufficient information.

Considering the conditions that five points are completely observed on an edge, and that the LRF does not collide with the robot manipulating the peg, the possible position of LRF is on the circle. The circle's center is located at the vertex at the intersection of the two neighboring edges. The laser plane is pointing toward the vertex. The position of the sensor is measured as the angle (ϕ) from one of the edge. Fig. 21 shows a successful peg-in-hole operation.

5.3.2 Verifying the Accuracy of the Predicted Success Probability

We verify the accuracy of the predicted success probability using the actual success ratio in the following two sets of the objects:

Case (a): The cross-section of the peg is a square of 19.05 [*mm*] × 19.05 [*mm*]. The clearance ratio of the hole is 0.043. The success region of this operation is depicted in Fig. 16a.
Case (b): The cross-section of the peg is a rectangle of 38.1 [*mm*] × 19.05 [*mm*]. The clearance ratio of the hole is 0.043. The success region of this operation is depicted in Fig. 16b.

In each case, the observation angle, ϕ, is sampled every 10 degrees; at each angle, the insertion operations were performed 50 times to obtain the ratio between success and failure.

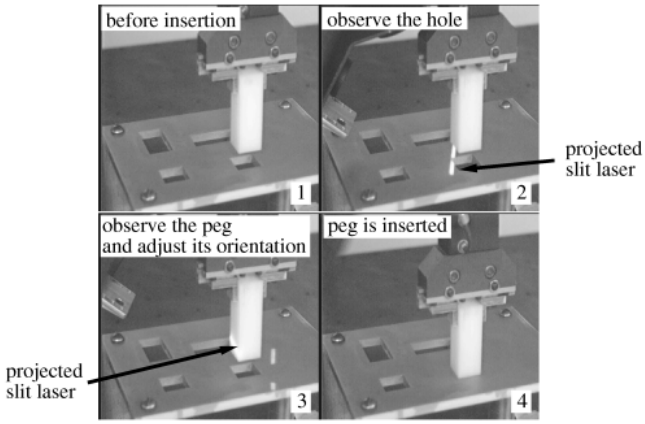


Fig. 21. A successful peg-in-hole operation.

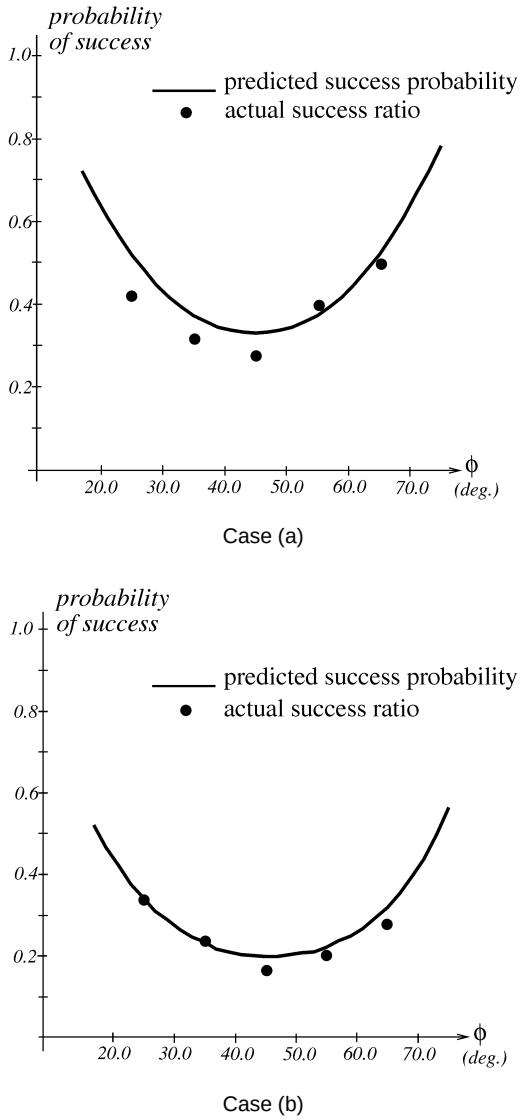


Fig. 22. Comparison of the predicted success probability with the actual success ratio in the peg-in-hole operation.

Fig. 22 shows the verification results; in both cases, the results given by actual performances (dots) coincide with the predicted success probability (solid curves).

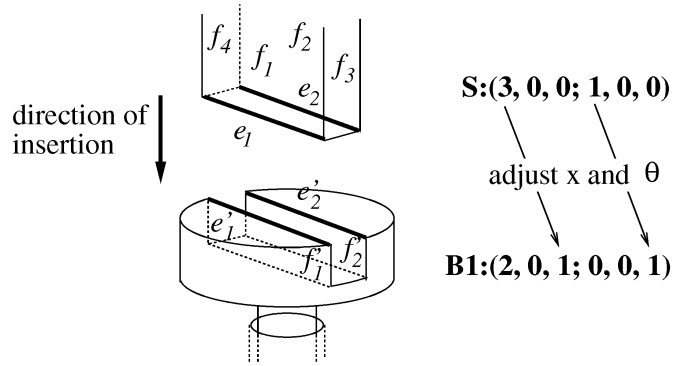


Fig. 23. Contact state analysis of putting a screwdriver on a screw.

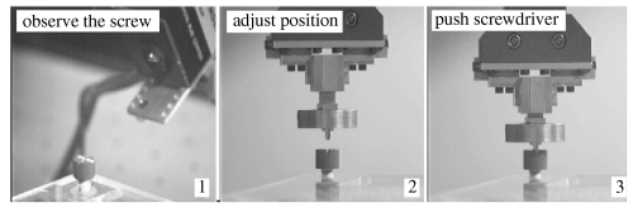


Fig. 24. The screwdriver was successfully inserted into the slot of the screw head.

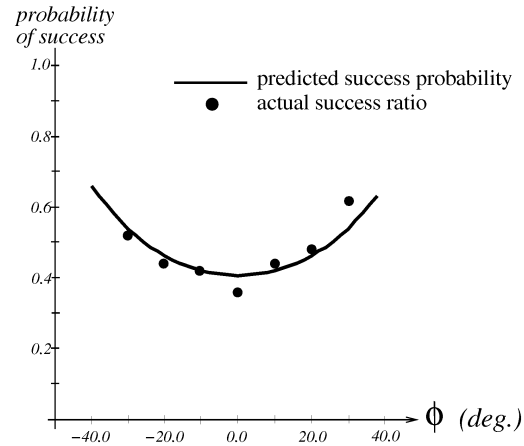


Fig. 25. Comparison of the predicted success probability with the actual success ratio in the screwdriver-bolt operation. The angle ϕ indicates the difference of the directions of the LRF and the slot.

5.4 Putting Screwdriver on Screw

This example considers the insertion of a screwdriver into the slot on a screw head as shown in Fig. 23. This operation belongs to group (c) in Fig. 10. During this operation, the critical DOF are one DOF for translation and one DOF for rotation. The face contacts to be achieved are $(f_1 - f'_1)$ and $(f_2 - f'_2)$. The candidates for observed features are f_1, f_2, f_3, and f_4 for the screwdriver, and e'_1 and e'_2 for the hole.

Due to the geometric constraints between arms for the screwdriver and for the LRF, the screwdriver and the screw could not be observed at once. Thus, the LRF observed only the screw because the positional uncertainty of the screw was much larger than that of the screwdriver. Edges e'_1 and

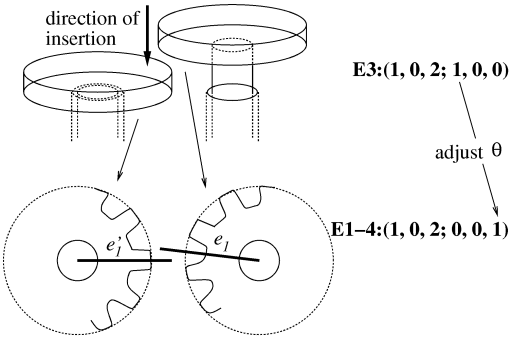


Fig. 26. Contact state analysis of gear mating.

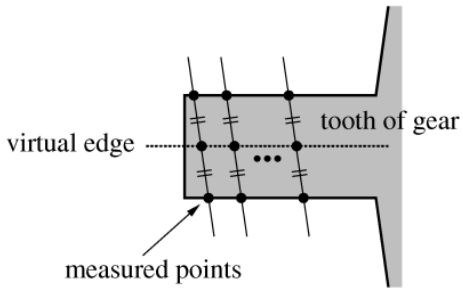


Fig. 27. Measuring the tooth position from edge positions.

e_2' of the screw were observed. Fig. 24 shows a successful operation of putting a screwdriver on a screw.

Fig. 25 shows the verification result of the predicted success probability with the actual success ratio. The predicted success probability coincides with the actual success ratio.

5.5 Gear Mating

A gear-mating operation, shown in Fig. 26, belongs to group (e) in Fig. 10. In this operation, a priori knowledge about how gears are mated is necessary because there are many potential matches between gear teeth. First, two virtual edges e_1 and e_1' are generated; one edge is placed on the center of the nearest tooth to the line connecting two gear centers; another edge is placed on the center of the nearest gap to the line. Then, the orientation of the inserted gear is adjusted so that these two virtual edges are aligned.

The position of a virtual edge is calculated from the position of the edges on the tooth (or gap) on which the virtual edge is set (see Fig. 27). Assuming that the shape of the tooth is almost rectangular, the virtual edge is obtained by fitting a line to the center points of the edge point pairs.

Fig. 28 shows a successful gear-mating operation. Fig. 29 shows the comparison of the predicted success probability with the actual success ratio. The predicted success probability is consistent with the actual success ratio.

6 Concluding Remarks

We have described a method of systematically generating visual sensing strategies using the knowledge of the task to be performed. The analysis of face contact transitions decides

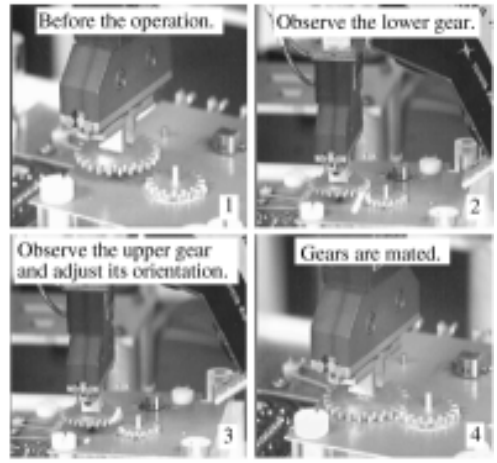


Fig. 28. The gears were successfully mated.

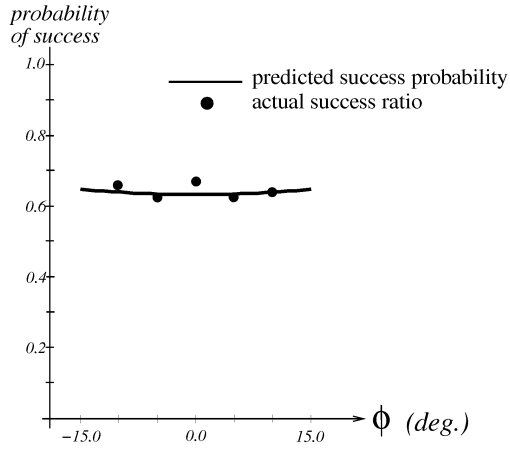


Fig. 29. Comparison of the predicted success probability with the actual success ratio in the gear-mating operation. The angle ϕ indicates the difference of the directions of the LRF and the line connecting two gear centers.

1) whether the current operation requires visual feedback and
2) which, if any, DOFs (critical DOF) is required to monitor during the operation through visual feedback.

Then, sensing primitives convert the critical DOF and the CAD models into a set of visual features to be observed. The final sensing strategy selected is the one with the highest predicted success probability among possible sensing strategies that observe the set of visual features. The proposed method is implemented using a line laser range finder as the sensor. The experiments are conducted to verify our evaluation method for choice of the optimal sensing strategy.

The features of the proposed method are summarized as follows:

1) the necessary visual information and the visual feature set to extract such information, which have been given in the previous works, are derived automatically from the task description;
2) the criterion using the predicted success probability to rank the set of sensing strategies is a general one and

is effective regardless of the shape of objects and the uncertainty models of the sensors used.

This paper has dealt with the assembly operations in which only face contacts are allowed among possible contacts between objects. Assembly tasks that require full six DOF motions cannot be analyzed using only surface normals. Thus, we are now reformulating the contact-state analysis using the theory of polyhedral convex cones [21]; the new analysis covers face, point, and line contacts and can, in principle, handle general six DOF motions. The arguments on necessary visual information in Section 2.2 could be applied to the result of this new analysis.

If point contacts and/or line contacts are allowed, most assembly operations can be achieved with only force information [9]. Even in such a case, visual information will be useful in, for example, reducing the number of motion steps [11]. A future work is to develop a method of coordinating vision and force information for robust and efficient assembly task execution.

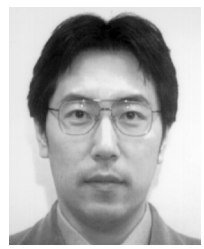
Another future work is to apply the proposed method to other sensors such as stereo vision.

ACKNOWLEDGMENTS

The authors would like to thank Marie Elm and Sing Bing Kang for helpful comments on the draft of this paper and Masato Kawade for the experimental setup. They would also like to thank the members of the Task-Oriented Vision Laboratory, The Robotics Institute, Carnegie Mellon University, for fruitful discussions.

REFERENCES

[1] J. Aloimonos, "Purposive and Qualitative Active Vision," *Proc. Image Understanding Workshop*, pp. 816–828, 1990.
[2] Y. Aloimonos, ed., *Active Perception*. Hilldale, N.J.: Lawrence Erlbaum Associates, Inc., 1993.
[3] N. Ayache and O.D. Faugeras, "Maintaining Representations of the Environment of a Mobile Robot," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 804–819, 1989.
[4] R. Bajcsy, "Active Perception," *Proc. IEEE*, vol. 76, no. 8, 1988.
[5] D.H. Ballard, "Reference Frames for Animate Vision," *Proc. Int'l Joint Conf. Artificial Intelligence*, pp. 1,635–1,641, 1989.
[6] L. Birnbaum, M. Brand, and P. Cooper, "Looking for Trouble: Using Causal Semantics to Direct Focus of Attention," *Proc. Fourth Int'l Conf. Computer Vision*, pp. 49–56, 1993.
[7] C.K. Cowan, "Automatic Camera and Light-Source Placement Using CAD Models," *IEEE Workshop on Directions in Automated CAD-Based Vision*, pp. 22–31, 1991.
[8] K.D. Gremban and K. Ikeuchi, "Planning Multiple Observations for Object Recognition," *Int'l J. Computer Vision*, vol. 12, no. 2, pp. 137–172, 1994.
[9] S. Hirai and H. Asada, "Kinematics and Statics of Manipulation Using the Theory of Polyhedral Convex Cones," *Int'l J. Robotics Research*, vol. 12, no. 5, pp. 434–447, 1993.
[10] I. D. Horswill, *Specialization for Perceptual Processes*, PhD thesis, Massachusetts Inst. of Technology, 1993.
[11] S.A. Hutchinson, "Exploiting Visual Constraints in Robot Motion Planning," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,722–1,727, 1991.
[12] S.A. Hutchinson and A.C. Kak, "Planning Sensing Strategies in a Robot Work Cell With Multi-Sensor Capabilities," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 765–783, 1989.
[13] K. Ikeuchi and M. Hebert, "Task Oriented Vision," Technical Report CMU-CS-91-163, School of Computer Science, Carnegie Mellon Univ., July 1991.
[14] K. Ikeuchi and T. Kanade, "Towards Automatic Generation of Object Recognition Program," *Proc. IEEE*, vol. 76, no. 8, pp. 1,016–1,035, 1987.
[15] K. Ikeuchi and T. Suehiro, "Toward an Assembly Plan From Observation Part I: Task Recognition With Polyhedral Objects," *IEEE Trans. Robotics and Automation*, vol. 10, no. 3, pp. 368–385, 1994.
[16] K. Kemmotsu and T. Kanade, "Sensor Placement Design for Object Pose Determination With Three Light-Stripe Range Finders," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,357–1,364, San Diego, Calif., 1994.
[17] Y. Kuniyoshi and H. Inoue, "Qualitative Understanding of Ongoing Human Action Sequences," *Proc. 13th Int'l Joint Conf. Artificial Intelligence*, pp. 1,600–1,609, Chambéry, France, 1993.
[18] T. Lozano-Pérez, M.T. Mason, and R.H. Taylor, "Automatic Synthesis of Fine Motion Strategies for Robots," *Int'l J. Robotics Research*, vol. 3, no. 1, pp. 3–24, 1984.
[19] M.T. Mason, "Compliance and Force Control for Computer Controlled Manipulation," *IEEE Trans. System, Man, and Cybernetics*, vol. 11, no. 6, pp. 418–432, 1981.
[20] O. Ozeki, K. Higuchi, and S. Yamamoto, "Automated Dimension Inspection System for Automotive Plastic Parts With a Laser Probe," *Proc. Robots 12 and Vision '88 Conf.*, pp. 5-51–5-60, Detroit, Mich., 1988.
[21] G.V. Paul and K. Ikeuchi, "Partitioning Contact-State Space Using the Theory of Polyhedral Convex Cones," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 421–426, 1995.
[22] R.D. Rimey, "Control of Selective Perception Using Bayes Nets and Decision Theory," Technical Report 468, Computer Science Dept., Univ. of Rochester, Dec. 1993.
[23] S. Sakane and T. Sato, "Automatic Planning of Light Source and Camera Placement for an Active Photometric Stereo System," *IEEE Int'l Conf. Robotics and Automation*, pp. 1,080–1,087, 1991.
[24] V. Scheinman, "A Multiple Robot Vision Guided Assembly System," R. Bolles and B. Roth, eds., *Robotics Research 4*. Cambridge, Mass.: MIT Press, 1987.
[25] A.J. Spyridi and A.A.G. Requicha, "Automatic Programming of Coordinate Measuring Machines," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,107–1,112, San Diego, Calif., 1994.
[26] K.A. Tarabanis, P.K. Allen, and R.Y. Tsai, "A Survey of Sensor Planning in Computer Vision," *IEEE Trans. Robotics and Automation*, vol. 11, no. 1, pp. 86–103, 1995.
[27] K.A. Tarabanis, R.Y. Tsai, and P.K. Allen, "The MVP Sensor Planning System for Robotic Vision Tasks," *IEEE Trans. Robotics and Automation*, vol. 11, no. 1, pp. 72–85, 1995.
[28] J.K. Tsotsos, "The Complexity of Perceptual Search Tasks," *Proc. 11th Int'l Joint Conf. Artificial Intelligence*, pp. 1,571–1,577, 1989.
[29] P. Whaite and F.P. Ferrie, "From Uncertainty to Visual Exploration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 72–85, Oct. 1991.
[30] C.C. Yang, M.M. Marefat, and R.L. Kashyap, "Active Visual Inspection Based on CAD Models," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,120–1,125, San Diego, Calif., 1994.
[31] H. Zhang, "Optimal Sensor Placement," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 1,825–1,830, 1992.



Jun Miura received the BEng degree in mechanical engineering in 1984 and the MEng and the DrEng degrees in information engineering in 1986 and 1989, respectively, all from the University of Tokyo, Tokyo, Japan. Currently, he is a research associate of the Department of Computer-Controlled Mechanical Systems, Osaka University, Suita, Japan. From March 1994 to February 1995, he was a visiting scientist at the Computer Science Department, Carnegie Mellon University, Pittsburgh, Pa., USA. His research interests include robotics, artificial intelligence, and computer vision. Dr. Miura received the Best Paper Award from the Robotics Society of Japan in 1997.



Katsushi Ikeuchi received the BEng degree in mechanical engineering from Kyoto University, Kyoto, Japan, in 1973 and the PhD degree in information engineering from the University of Tokyo, Tokyo, Japan, in 1978. After working at the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology, the Electrotechnical Laboratory of the Ministry of International Trade and Industries, and the School of Computer Science, Carnegie Mellon University, he joined the University of Tokyo in 1996. He is currently a professor of electrical engineering at the Institute of Industrial Science, the University of Tokyo, Tokyo, Japan. Dr. Ikeuchi developed the "smoothness constraint," a constraint to force neighboring points to have similar surface orientations. Recently, his main focus is to develop vision techniques that enable to reduce programming efforts. He has been the general or program chair of the 1995 Intelligent Robotics and Systems conference and the 1996 International Conference on Computer Vision and Pattern Recognition. He served on the program committees of several international conferences. He is on the editorial board of the *IEEE Transactions on Robotics and Automation, International Journal of Computer Vision, Journal of Manufacturing Systems,* and *Optical Society of America*. He has received several awards, including the David Marr Prize and an IEEE outstanding paper award. In addition, his 1992 paper, "Numerical Shape From Shading and Occluding Boundaries," was selected as one of the most influential papers to have appeared in *Artificial Intelligence Journal* in the past 10 years. He is a Fellow of the IEEE.