Proc. 1995 IEEE Int. Conf. on Robotics and Automation
pp.1912-1918, Nagoya, Japan, May 1995

Generating Visual Sensing Strategies in Assembly Tasks*

Jun Miura

Department of Mechanical Engineering
for Computer-Controlled Machinery
Osaka University
jun@cv.ccm.eng.osaka-u.ac.jp

Katsushi Ikeuchi

School of Computer Science
Carnegie Mellon University
ki@cs.cmu.edu

Abstract

It is generally very difficult, if not impossible, for a robot to perform fine manipulation tasks without the benefit of some form of sensory feedback during actual task execution. As a result, sensing planning is an important component in assembly task planning. This paper describes a method of generating visual sensing strategies based on knowledge of the task to be performed. The generation of the appropriate visual sensing strategy entails knowing what *information to extract and* where *to get it. This is facilitated by the knowledge of the task, which describes how objects are assembled. This knowledge, coupled with known sensor modeling, results in an abstract template of sensing strategy called the* sensing task model. *By instantiating the appropriate sensing task model at planning time, the sensing strategy is efficiently generated. Our method has been implemented using a laser range finder as the sensor. Experimental results involving typical assembly tasks show the feasibility of the method.*

1 Introduction

We have been developing a novel robot programming system, the APO (Assembly Plan from Observation) system [4]. 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.

Uncertainty of robot motion and errors in object modeling result in inconsistency between the task description and the world in which the robot operates. Such inconsistency, which may cause object-manipulation collision, for example, has been ignored in the previous system. This paper proposes a method of automatically generating sensing strategies for visual feedback in order to resolve the inconsistency.

In assembly tasks, the sensing planning problem [2] has been mainly concentrated on the *sensor placement problem*, that is, the sensing condition is determined which satisfies several requirements on imaging such as resolution, field of view, focus and visibility in a static [1][7][10] or in a dynamic [6] environment. Features to be observed are usually given beforehand, and are not automatically selected from a task or problem specification.

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 [8] or are selected from the specification of entities to be measured through given knowledge of mapping from measurable entities of an object to features to be observed [11].

In vision-guided operations, visual sensors should be strategically placed to extract relevant information for proper task execution. To determine *what* information is required and *where* to get it, 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.

In this paper, we propose a method of generating visual sensing strategy in assembly tasks by analyzing the task description. In assembly operations, degrees of freedom of assembled objects are gradually constrained. Thus, specific degrees of freedom of the currently manipulated objects need to be observed in each assembly operation. The description of the current operation indicates the degrees of freedom that should be measured. The description also provides a set of candidates of features to be observed. Based on such information, coupled with knowledge of the relationships between observed features and degrees of freedom to be measured, the system can then automatically generate a set of features by observing which the necessary degrees of freedom are measured.

We introduce a new representation called the *sensing task model*, which is a template of necessary sensing strategy for each assembly operation, for efficient generation of sensing strategy. The proposed method is implemented for a laser range finder as the sensor. Experimental results of typical assembly operations are described.

*This research was sponsored by the Advanced Research Projects Agency under the Department of the Army, Army Research Office under grant number DAAH04-94-G-0006.

2 Determining What Visual Information is Necessary

Visual information can be effectively used in certain types of assembly operations, while other types can be performed without visual information if the robot is capable of compliant motion. In this section, we first describe the task analysis based on face contact relations between objects. Then, using the result of the analysis, we explain how to determine what visual information is necessary for each assembly operation.

2.1 Task Analysis Based on Face Contact State

We analyze a state of the environment in terms of face contacts between object surfaces [4]. This analysis first deals with the case where polygonal objects perform only translational motions, and is extended later (see Section 2.4). We assume that each assembly operation (i.e., transition of contact state) involves one manipulated object, manipulated by a robot for the current operation, and several stationary environmental objects which have face contacts with the manipulated object. We also assume that the goal of each assembly operation is to establish the required face contact state.

2.2 Representation of Face Contact States

Let us suppose a surface patch of the manipulated object have a face contact to a surface patch of an environmental object. This surface contact pair constrains the manipulated object's possible translation motion by:

$$\boldsymbol{N} \cdot \Delta \boldsymbol{T} \geq 0,$$

where $\Delta \boldsymbol{T}$ denotes possible translational motion vectors of the manipulated object and $\boldsymbol{N}$ denotes the normal direction of an environmental surface patch.

We use points on the Gaussian sphere to specify both a constraint vector and all possible translation vectors. Each 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 point to denote the vector.

The constraint from a patch pair defines several regions in the Gaussian sphere (see Fig.1). Assuming that the normal, $\boldsymbol{N}$, points to the north pole of the Gaussian sphere without loss of generality, the northern hemisphere corresponds to possible motion directions; the southern hemisphere corresponds to prohibited motion directions.

In Fig. 1, motions of the directions corresponding to the boundary of the southern hemisphere (the equator) maintain the current face contact state. The degrees of freedom of the maintaining the contact state (*maintaining DOF*) is two. Motions of the directions corresponding to the inside of the detaching hemisphere break the contact state, and is referred to as the *detaching* motion. A pure detaching motion is the detaching motion which does not contain any maintaining motion component. The pure detaching motion in Fig. 1 is along the constraint normal $\boldsymbol{N}$; its degrees of freedom (*detaching DOF*) is one.

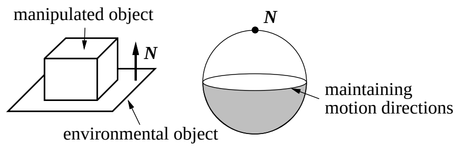

Fig. 1: Constraint depicted on the Gaussian sphere.

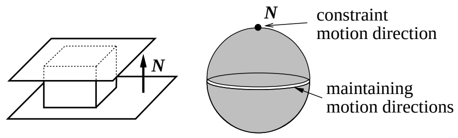

Fig. 2: A bidirectional constraint.

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. One direction along the surface normals is completely constrained; the degrees of freedom of the constraint directions (*constraining DOF*) is one.

We can specify a face contact state by using a triplet of maintaining, detaching, and constraining DOFs. Using this triplet, for example, the states of Figs. 1 and 2 are represented as $(2, 1, 0)$ and $(2, 0, 1)$, respectively. Each assembly operation is considered as a transition from one contact state to another. In this analysis, we extracted ten possible contact states and thirteen possible transitions [4].

2.3 Determining What Visual Information is Necessary

An assembly operation always increases constraints on some degrees of freedom of the manipulated object. This increase of constraint is classified into three cases: from maintaining DOF to detaching DOF, from detaching DOF to constraining DOF, and from maintaining DOF to constraining DOF. Fig. 3 shows typical situations corresponding to the three cases.

Let us examine how the type of the degree of freedom for horizontal motion changes in these cases. In case (a), the degree of freedom changes from maintaining DOF to detaching DOF. Since the approaching direction of the block is parallel to the direction of the pure detaching motion at the final state (i.e., the normal vector of the wall), this operation is realized by moving the block until the face contact occurs. Thus, this operation can be performed by compliant motion

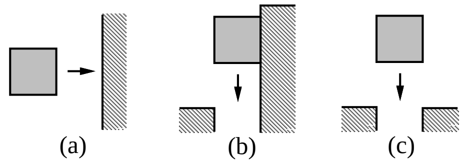

Fig. 3: Three typical cases of increase of constraint. Types and transitions of the triplet are:
(a): maintaining $\rightarrow$ detaching $((3,0,0) \rightarrow (2,1,0))$.
(b): detaching $\rightarrow$ constraining $((2,1,0) \rightarrow (2,0,1))$.
(c): maintaining $\rightarrow$ constraining $((3,0,0) \rightarrow (2,0,1))$.

without visual information. In case (b), the degree of freedom changes from detaching DOF to constraining DOF. Although the horizontal degrees of freedom is constrained at the final state, visual information is also unnecessary because the desired horizontal position can be kept by maintaining the contact between the block and the right wall.

In case (c), the degree of freedom changes from maintaining DOF to constraining DOF. The horizontal position of the block needs to be adjusted with visual information before mating so that both the left and the right face contact are achieved simultaneously. Since there is no contact before mating, force information cannot be used.[1]

To summarize, if a degree of freedom becomes constraining DOF from maintaining DOF, that degree of freedom should be observed. By applying this theory to thirteen possible transitions, four transitions were found to require visual information [5].

2.4 Extension of Analysis of Face Contact State

In this paper, we extend the previous contact state analysis to the case where an object can be composed of planar or cylindrical surfaces, and where, in addition to three translational degrees of freedom, one rotational degrees of freedom is allowed in one face contact state transition. Fig. 4 shows a typical object used in the analysis. We limit the contact states of a cylindrical surface to the three cases where (from left to right) no contact occurs, contact occurs on half of its surface, and contact occurs on all of its surface; these cases correspond to maintaining DOF, detaching DOF, and constraining DOF for the translational motion perpendicular to the principal axis of the cylindrical surface. In this extended analysis, we represent a face contact state by a sextuplet of DOF, which is composed of two triplets for translational DOFs and for rotational DOFs. This analysis can cover a relatively large number of actual assembly operations. Refer to [5] for details.

[1] A sophisticated force control-based manipulation strategy may be employed to perform this kind of assembly operation without visual feedback [9]. Even in such a case, reducing errors by visual information would be useful.

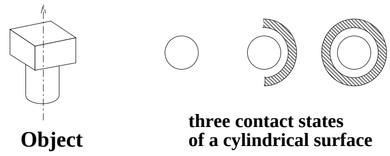

Fig. 4: The object used in the extended face contact state analysis.

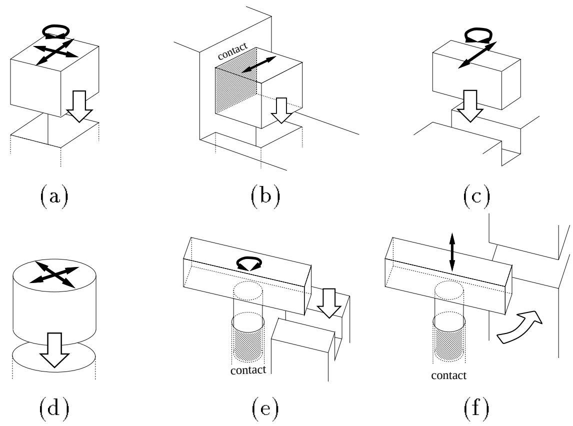

Fig. 5: Transition groups which need visual information. Thick arrows indicate the direction of movement. Thin arrows indicate degrees of freedom to be adjusted by use of visual information. The transitions of the sextuplet are:
(a): $(3,0,0;1,0,0) \rightarrow (1,0,2;0,0,1)$.
(b): $(2,1,0;0,0,1) \rightarrow (1,0,2;0,0,1)$.
(c): $(3,0,0;1,0,0) \rightarrow (2,0,1;0,0,1)$.
(d): $(3,0,0;1,0,0) \rightarrow (1,0,2;1,0,0)$.
(e): $(1,0,2;1,0,0) \rightarrow (1,0,2;0,0,1)$.
(f): $(1,0,2;1,0,0) \rightarrow (0,0,3;1,0,0)$.

The theory to detecting necessary visual information is also applicable to rotational motions because an infinitesimal rotational motion just before the state transition is considered as a translational motion. By applying the theory to the result of the extended contact state analysis, 19 out of 85 transitions were found to require visual information. Further examination of transitions in terms of the change of the sextuplet classified these 19 transitions into six groups. Typical situations for the groups are depicted in Fig. 5.

3 Selection of Features to be Observed

In each vision-guided assembly operation, a relevant set of features needs to be selected so that necessary degrees of freedom of the assembled objects are observed.

3.1 Sensing Primitive

To solve the feature selection problem, we introduce a concept of *sensing primitive*. Sensing primitive is an

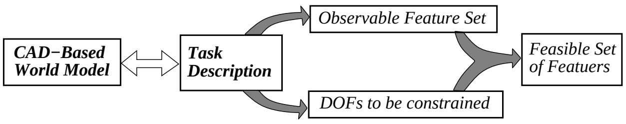

Fig. 6: Selection of features to be observed using task description and sensing primitives.

abstract sensing procedure, which describes the relationship between an observable feature and degrees of freedom to be measured. For each primitive visual feature, such as a straight edge of a polyhedron, one sensing primitive is prepared. The repertoire of sensing primitives is generated in advance by enumerating possible geometric features that could appear in the assembly task under consideration, and that are observable by the sensor used.

3.2 Feature Selection Process

The feature selection is performed as follows (see Fig. 6). From the task description, the degrees of freedom to be constrained by the assembly operation can be obtained. On the other hand, an observable feature set comes from the face contact information in the task description. By consulting prepared *sensing primitives*, a feasible set of features is selected. Necessary geometric information in this selection is retrieved from a CAD-based world model.[2]

For example, let us consider cases (a) and (b) in Fig. 5. Suppose that we are localizing the hole by observing its edges, and that we prepare a sensing primitive to observe a straight edge; this sensing primitive can measure the edge's position perpendicular to that edge. Candidate features are four edges of a hole. In case (b), an edge of the hole perpendicular to the movable direction of the block provides sufficient information for localization, while a pair of neighboring edges should be observed in case (a).

4 Sensing Strategy Generation using Sensing Task Models

We introduce a new representation termed the *sensing task model* for efficient generation of sensing strategy. This section first explains the APO framework before describing how the sensing task models are used in the APO system. Fig. 7 illustrates the outline of sensing strategy generation.

4.1 Task Model and Sensor Model

The description of a task is stored in a structure called an *abstract task model*. An abstract task model associates a state transition with an assembly task which causes that transition. Each task model has slots for necessary information for performing the operation by a robot, such as assembled objects and geometric relations to be achieved. In addition, the task model contains a robot motion macro and parameters to expand the macro, such as grasping position, departing position and approaching position. An *instantiated task model*, whose slots have actual values, is generated by observing an assembly operation performed by a human in the APO system. Or, an assembly planner could generate instantiated task models from the task specification.

The *sensor model* [3] describes knowledge about a sensor such as features which are observable with the sensor, range of the sensor (distance and field of view), and sensor data uncertainty.

[2] The object recognizer determines each object configuration in the real world and generates a CAD-based world model.

4.2 Sensing Task Model

Sensing strategies are dependent upon the sensors used and assembly operations to be performed. An *abstract sensing task model* is generated for a sensor and an assembly operation (or a group of assembly operations) from the sensor model and abstract task models. An abstract sensing task model contains the following information:

- *Transition of the sextuplet*
- *Which degrees of freedom to observe.*
- *Feature set to observe*
- *feasible sensor position set* to observe it.

Some of the above information are dependent upon geometric values (shape and size) of the objects involved in each operation. In order to efficiently generate visual sensing strategies, we enumerate in advance operations which involve objects of typical shapes such as rectangular parallelepiped or cylinder, and describe the above information in a parameterized form for those operations. Parameters (i.e., the size of objects involved in each operation) are instantiated at planning time by referring to the instantiated task model. Fig. 8 shows a parameterized abstract sensing task model for the operation of inserting a peg with a circular cross-section into a hole.

For those operations that do not have corresponding parameterized information, the necessary information is generated from scratch, that is, the system determines degrees of freedom to be observed, selects

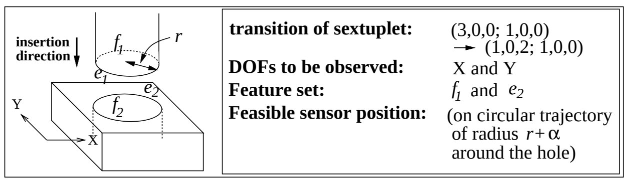

Fig. 8: An example of parameterized abstract sensing task model. f_1 and f_2 indicate faces; e_1 and e_2 indicate holes; r is the radius of the cross-section of the peg; α is determined so that there is no collision of the sensor.

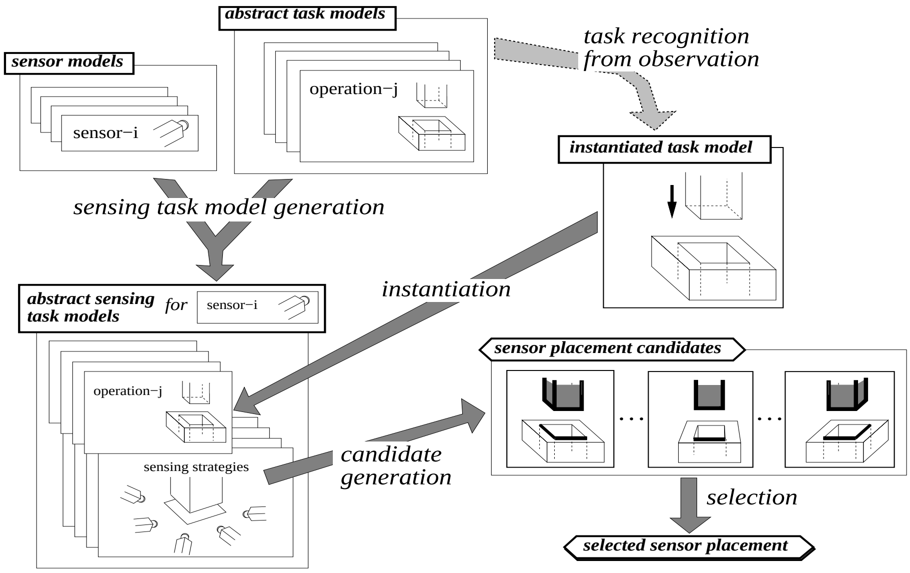

Fig. 7: Framework of sensing strategy generation in the APO system.

feasible feature, and determines a feasible sensor position set by considering visibility and detectability of the features.

4.3 Process of Sensing Task Model Generation

Before planning, abstract sensing task models are generated for a sensor model and a set of abstract task models using the following steps:

1. Collect and classify state transitions in which visual information should be used. This step is based on the analysis of transitions of contact states (see section 2).
2. For each state transition groups, enumerate possible shapes of objects, for which parameterized abstract sensing models are prepared. A parameterized abstract sensing model is generated in the following steps:
   (a) Determine a feasible sets of features to observe.
   (b) Determine feasible sensor positions the above set by considering visibility and detectability of the features.

4.4 Automatic Sensing Strategy Generation using Sensing Task Model

At planning time, given an instantiated task model and the sensing task models, sensing strategy for the operation is automatically generated by the following steps (see Fig. 7):

1. *Instantiation.* Select the appropriate abstract sensing task model for the current assembly operation, and instantiate it using actual geometric values.
2. *Candidate generation.* From the instantiated sensing task model, possible sensor positions are generated. In this step, the possibility of collision between the sensor and the environment including other robots is examined.
3. *Selection.* Select one of feasible sensing strategies using a certain evaluation function.

5 Implementation of the Method using Laser Range Finder

The proposed method has been implemented using a line laser range finder (LRF). The LRF emits slit laser, detects highlighted portion of the object by a TV camera (see Fig. 9), and obtains a line of 3D measurement. The LRF is attached to a manipulator with four degrees of freedom, three degrees of freedom for translation and one for rotation around the vertical axis.

Every assembly operation that requires visual information is a kind of "peg-in-hole" operation. The location of a peg is measured by observing its side faces; that of hole is measured by observing several 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 face.

We use a sensing strategy as shown in Fig. 10; data for one assembly operation are collected at several position by moving the LRF in parallel with the

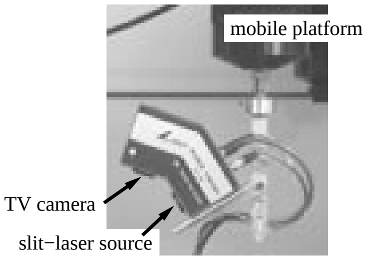

Fig. 9: Laser range finder.

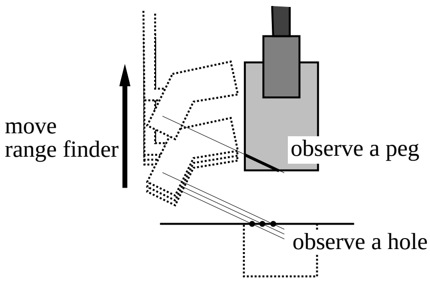

Fig. 10: A strategy for observing a peg and a hole.

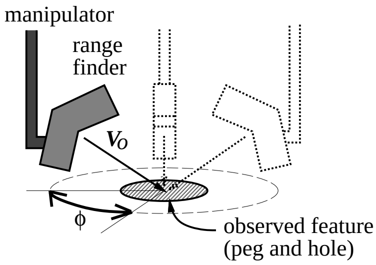

Fig. 11: Candidate positions.

insertion direction because the relative displacement on the plane perpendicular to the insertion direction is important for the position adjustment of the peg. We also control the position of the LRF so that each measured point is kept within a certain area of the slit laser. Thus, the only parameter that specifies the position of the LRF is the angle ϕ between the direction of the laser and some axis on the horizontal plane (see Fig. 11).

6 Experimental Results

6.1 Assembly Operation with Visual Feedback

The process of an actual vision-guided assembly operation is as follows. First, a peg is moved by a manipulator to the position just before a hole. Then, the LRF is placed to the planned position, and measures the position of the hole and the peg. If the estimated maximum error in the relative position between the peg and the hole is within the predetermined clearance, the peg is inserted. Otherwise, the peg position is adjusted and the position is re-observed. This final step is repeated until the relative position becomes satisfactory, and then the peg is inserted.

6.2 Putting Screwdriver on Bolt

Fig. 12 shows an operation to insert the tip of a screwdriver into the slot on a bolt head. This operation belongs to group (c) in Fig. 5. By this operation, two degrees of freedom are constrained. 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.

In this case, because of the geometric constraints between manipulators, the screwdriver and the bolt could not be observed at once. Thus, the LRF observed only the bolt since the positional uncertainty of the bolt is much larger than that of the screwdriver. Thus, by observing edges e'_1 and e'_2 of the bolt, the operation was completed as shown in Fig. 13.

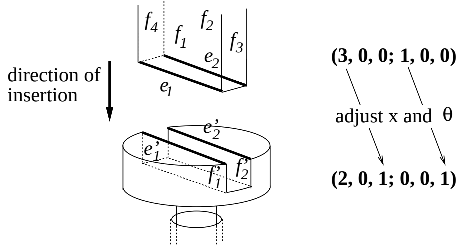

Fig. 12: Contact state analysis of the operation of putting a screwdriver on a bolt. Transition of the sextuplet of DOFs (see section 2) is also indicated.

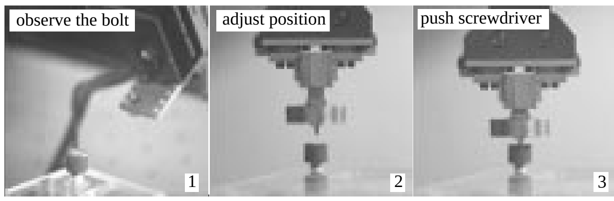

Fig. 13: The screwdriver was successfully inserted into the slot of the bolt head.

6.3 Gear Mating

Figure 14 shows a gear-mating operation. This transition belongs to group (e) in Fig. 5. In this transition, *a priori* knowledge about how cogs of gears are mated is necessary because there are many potential matches between cogs of gears. First, two virtual edges e_1 and e'_1 are generated; one edge is placed on the center of the nearest cog (or gap) to the line connecting two gear centers; another edge is placed on the center of the nearest gap (or cog) to the line. Then, the orientation of the inserted gear is adjusted so that these two virtual edges are aligned. Fig. 15 shows a successful gear mating operation.

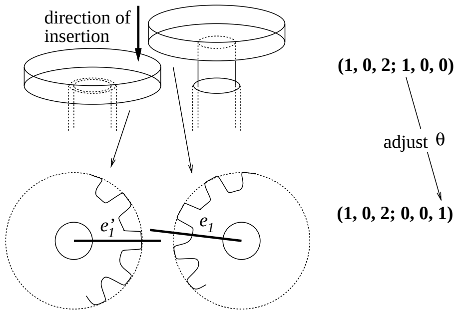

Fig. 14: Contact state analysis of gear mating operation.

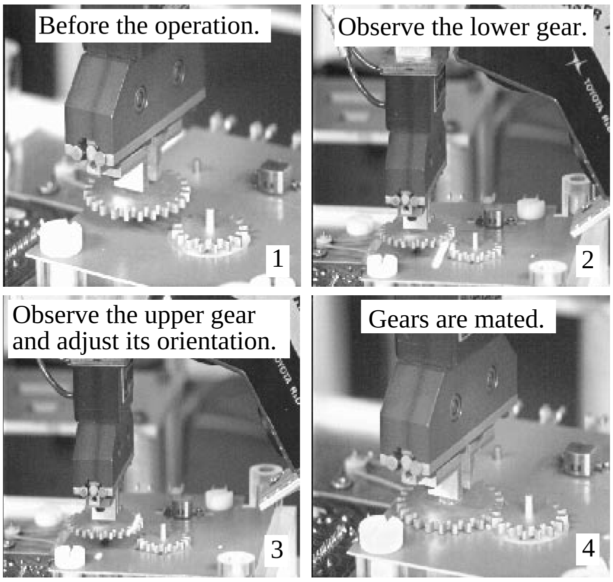

Fig. 15: The gears were successfully mated.

7 Conclusion

We propose a method of generating visual sensing strategies based on knowledge of the task to be performed. We analyze the task in terms of the transition of face contacts between object surfaces, and derive groups of assembly operations which require visual feedback. We introduce the notion of the *sensing task model*, which is a template of sensing strategy generated for each group of operations. By instantiating the appropriate sensing task at planning time, the sensing strategy is efficiently generated. We have implemented our method using a laser range finder as the sensor. Experimental results involving typical assembly tasks show the feasibility of the method.

At present, sensing task models are generated partially manually. Automatic generation of the sensing task models from the task description and the sensor model is part of our future work. This work would make it easier to apply our method to other sensors such as stereo vision.

Acknowledgement

The authors would like to thank Sing Bing Kang for helpful comments on the draft of this paper and Masato Kawade for the experimental setup.

This research was conducted while the first author was with Carnegie Mellon University. His stay was supported by Telecommunications Advancement Foundation, Tokyo, Japan.

References

[1] C.K. Cowan. Automatic Camera and Light-Source Placement Using CAD Models. In *IEEE Workshop on Directions in Automated CAD-Based Vision*, pages 22–31, 1991.

[2] S.A. Hutchinson and A.C. Kak. Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities. *IEEE Trans. on Robotics and Automat.*, RA-5(6):765–783, 1989.

[3] K. Ikeuchi and T. Kanade. Modeling Sensors: Toward Automatic Generation of Object Recognition Program. *Computer Vision, Graphics, and Image Processing*, 48:50–79, 1989.

[4] K. Ikeuchi and T. Suehiro. Toward an Assembly Plan from Observation Part I: Task Recognition With Polyhedral Objects. *IEEE Trans. on Robotics and Automat.*, 10(3):368–385, 1994.

[5] J. Miura and K. Ikeuchi. Task-Oriented Generation of Visual Sensing Strategies in Assembly Tasks. Technical Report CMU-CS-95-116, School of Computer Science, Carnegie Mellon University, February 1995.

[6] B. Nelson and P.K. Khosla. Integrating Sensor Placement and Visual Tracking Strategies. In *Proceedings of 1994 IEEE Int. Conf. on Robotics and Automat.*, pages 1351–1356, San Diego, CA, 1994.

[7] S. Sakane and T. Sato. Automatic Planning of Light Source and Camera Placement for an Active Photometric Stereo System. In *1991 IEEE Int'l Conf. on Robotics and Automat.*, pages 1080–1087, 1991.

[8] A.J. Spyridi and A.A.G. Requicha. Automatic Programming of Coordinate Measuring Machines. In *Proceedings of 1994 IEEE Int. Conf. on Robotics and Automat.*, pages 1107–1112, San Diego, CA, 1994.

[9] T. Suehiro and K. Takase. Representation and Control of Motion in Contact and Its Application to Assembly Tasks. In H. Miura and S. Arimoto, editors, *Robotics Research 5*, pages 367–374. The MIT Press, 1990.

[10] K. et al. Tarabanis. Satisfying the Resolution Constraint in the MVP Machine Vision Planning. In *1990 DARPA Image Understanding Workshop*, pages 850–860, 1990.

[11] C.C. Yang, M.M. Marefat, and R.L. Kashyap. Active Visual Inspection Based on CAD Models. In *Proceedings of 1994 IEEE Int. Conf. on Robotics and Automat.*, pages 1120–1125, San Diego, CA, 1994.