International Workshop on Artificial Intelligence for Industrial Applications 1988

KNOWLEDGE REPRESENTATION FOR IMAGE BASED ROBOT OPERATION

Tomoyuki Hamada, Kohji Kamejima and Ikuo Takeuchi

Mechanical Engineering Research Laboratory, Hitachi, Ltd.
502 Kandatsu, Tsuchiura, Ibaraki 300, Japan

Abstract: A knowledge representation method is proposed for understanding operator messages to generate motion of the robot mechanism. With this method, a generation process which activates a set of rewrite rules translates the operator's message to a sequence of symbolic motion modules, called motion schema. Further an instantiation process of the motion schema which binds schemata with geometric data stored in the work space model generates instant motion. Through an experimental study using a prototype system, certain aspects of the method were verified.

1. Introduction

Humans and robots have essential differences in their physical structure that affect specific work related capabilities. Robots, for example, are not hampered by biological limitations and thus are often required to work in space or in a radioactive environment. In these cases, robots must work in a physically isolated area away from humans, and yet must work under the control of humans, who are the motivators of the task. Thus smooth communication, i.e., co-possession of the information environment, between humans and robots is required. The situation is also similar in general factory robot applications. That is, robots whose physical composition is different from humans, should be physically isolated from factory workers. In addition, closely related information channels between humans and robots, called human access paths, should be prepared. This would greatly decrease the safety problems associated with robot application in factories.

One realization of the human access path, proposed in this paper, is an "image based operation" [1]. It has an intelligent process capable of accepting mental images which the operator formulates in his mind for the task. In this paper, we will describe the knowledge representation method for image based operation. In this method, three fundamental concepts are presented. First is motion schema which describes primitive motion idea in three elements; schema slots containing the skill required to obtain the object feature and two procedural descriptions, one generates instant motion trajectory while the other modifies an internal model of the real world. Second is the action network which accepts multi-dimensional operation demand and ambiguity in the operator's mental image features. Third is the work space model which accepts real world conditions and provides the task context with the motion schema and the action network similar to the blackboard system.

2. Concept of Image Based Operation

Generally each motion composing a task can be represented by words, such as "grasp," "attach" or "assemble." These words are the symbols bound to the concept structured for the meaning of each motion. This concept level utilizes the terms for thinking, planning and executing tasks used by humans. This motion concept level should be appropriate for the human access path for human-robot communication. We call this concept level "the mental image of the task" and we call the operation style supporting the mental image level communication "image based operation."

The words bound to the motion concept indicate a minor task goal. The motion concept entity, which achieves this task goal, is the motion subset. Therefore, the key technology for image based operations concerns the mechanism that specifies an instant motion trajectory relevant to the work space condition from the subset defined by a given task goal. We realized this mechanism through motion schema, action network and work space model.

The concept of image based operations is shown in Fig.1.

In image based operations, the operator gives the robot a command, called an action message, which consists of the word indicating the task goal and the objects to be handled. The operator expresses his intent with this action message. The action network translates it to a list of motion schema, and the motion schema defines the motion subset of each of its primitive task goals. The work space model captures real work space conditions and provides the task context that specifies the instant motion trajectory from the subset.

The work space model also has a view; a graphic display indicating work space conditions obtained by the model itself. The operator can directly specify the objects

88CH2529-6/88/0000-0417 $1.00 © 1988 IEEE

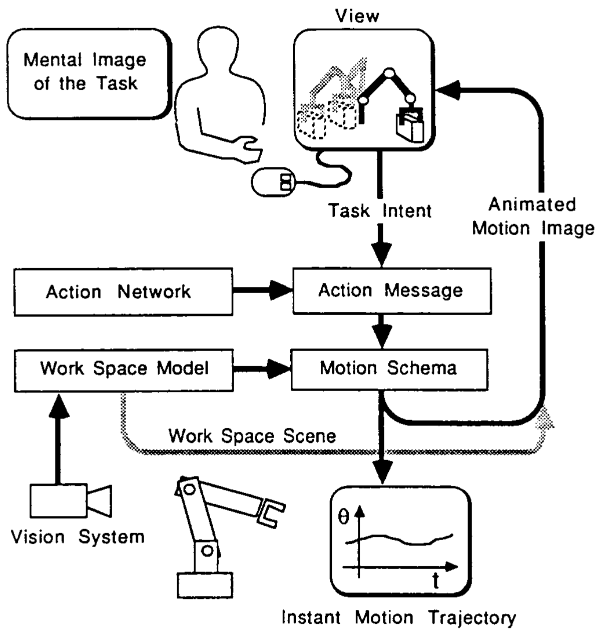

Fig. 1. Concept of image based robot operations.

to be handled through this view. Also with this view the robot displays the animated motion image decided for a given action message to the operator.

3. Considered Situation of the Robot Application

To clarify this study, we defined the robot application situation.

The environment is an indoor area where some industrial components are located. The geometric data concerning work space and components can be provided by a CAD system in a computer available format. The robot is required to assemble or disassemble those industrial components.

This assumption limits the motion control task of the robot to a geometric trajectory generation problem which does not contain the complicated servo parameter adaptation problems. A force control may be required for such tasks as driving screws or insertion. But in most cases, it would be limited to the decision problem of locations and orientation of parts to be attached. Connection and placement of each part changes dynamically. But the part shapes are clearly specified by their names. Therefore, the robot only has to comprehend the placement of the parts.

The command given by the operator is regarded as a brief task instruction which humans generally complete in one action, e.g., insert a wedge into the valve body or put a cap onto the body. The robot works automatically generating instant motion from this instruction. Considering the changing status of parts connections, the operator makes plans concerning the assembly process.

4. Motion Schema

The motions represented by the time variant arm joint angles or motions written by a robot language which contain absolute values of trajectory coordinates are the only instance of the motion subset specified by the task goal. Using those representations, each instant motion appears in different expressions depending on the placement of objects in the work space. The motion schema provides a uniform expression regardless of placement and it integrates instant motions into a unique symbol corresponding to the primitive task goal, i.e., the basic motion concept of humans.

The structure of the motion schema is shown in Fig.2. The principal mechanism of this integration is based on a cognitive psychological notion of human behavior [2]. According to this notion, when a person attempts to handle an object, he is supposed to recognize definite features of the object prior to the action. The information to be recognized is closely related to the intent of the action. The recognized information determines the actual instant motion trajectory. The information includes shape, type, placement and connection relation of each object and we call this information the work space context.

We realized this mechanism through the perception and execution parts of the motion schema. The perception part has some slots corresponding to the context required for the motion. When a motion schema is bound to an object, those slots are filled by the context values extracted from the object. After all slots are filled, the motion schema makes an instant motion trajectory based on the procedure written in the execution part. We call this process instantiation of the motion schema.

A motion schema for "grasp" is shown as an example in Fig.2. Slots GP,AP, W respectively mean grasp point, approach point, and grasp width of the target object. When the slot filling process is successful, an instant motion trajectory is generated by the procedure; move to AP, gripper open W, move to GP, and so on. When the slot filling process fails, the motion schema cannot be instantiated. This means that the motion is unexecutable in the current context of the work space. This situation may occur, for example, when the robot attempts to pick up a part which is firmly attached to another part or attempts to grasp parts that are too large to handle. With this instantiation mechanism, each motion schema can distinguish valid contexts in which to work. We refer to this as the context perception ability of the motion schema.

Execution of a motion makes some changes in the work space context. These changes are the side-effects

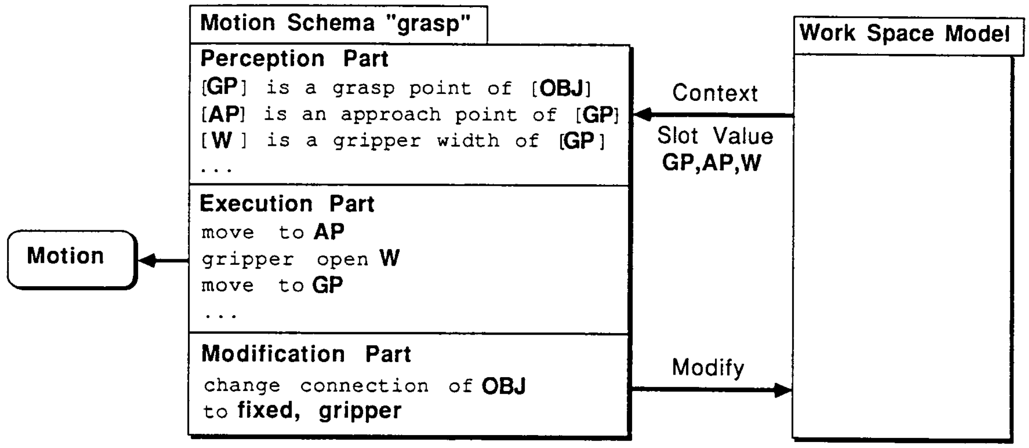

Fig. 2. Structure of the motion schema.

which the executor expects from the motion. Therefore, the context changes caused by the motion should also be integrated into the motion schema. The modification part of the motion schema contains the procedure which modifies the work space model to adjust to the expected changes. This procedure maintains the consistency of the work space model with the real world, and eliminates potential reconstruction of the work space model to correct gross miss-match caused by the motion.

An example of the motion schema written in Prolog language is shown in Fig.3. Using the motion schema, we can consistently represent motions based on the motion concept of humans.

5. Action Network

Once the motion schemata are prepared, generalized actions are described as a sequence of motion schemata. That is, a transition network which has the motion schemata as nodes can reasonably represent an action, and it generates an action sequence through the transitions between nodes. Then the operator's instruction is to specify the start symbol of the network by the action message. We call this network the action network. An example of the action network is shown in Fig.4.

The action network has two structural characteristics that accept the operator's mental image.

The conceptual level of the task goal based on the mental image is distributed into several classes. For examples, when a pipe joining task is requested, the individual bolt driving task, which is routine work, is not mentioned. On the other hand, the operator will want to carefully direct the bolt driving task in a valve assembly operation. To accept this multi-dimensional operation demand, the action network has a hierarchical structure. It allows each network node to be a start symbol of the other sub-network. This is the first characteristic of the action

```
grasp(OBJ):-
Perception Part
    not attached(OBJ,_),
    grasp_point(GP, OBJ),
    grasp_width(W, GP),
    approach_point(AP, GP),

Execution Part
    above(DTL, OBJ),
    location(P2,GP,OBJ),
    location(P1,AP,P2),
    location(P3,GP,DTL),
    with, gripper_open(W),
    move_to(P1),
    move_to(P2),
    with, gripper_close(W),
    stop,
    move_to(P3),

Modification Part
    get_i(GP,location,L),
    put(OBJ,connection,fixed,gripper),
    put(OBJ,location,value,L),
    make_instance_of(GP,P),
    clrfrm.
```

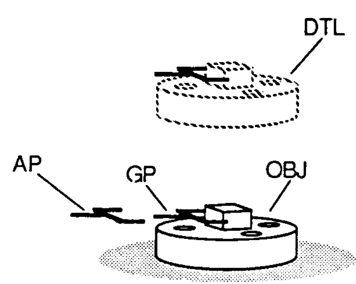

Fig. 3. Example of the motion schema "grasp" written in Prolog.

network.

Human conceptual hierarchy, however, is not clearly classified. There is an ambiguity in that an action message is translated to different lists of motion schemata depending on work space context. For example, an action "bind," which binds two objects, should be translated to different motion schemata, such as "fix with bolts," "insert" or "stick" according to the available binding method for the two objects. Therefore, the action network should have a structure to deal with this ambiguity.

This is accomplished by the context perception ability of the motion schema. The action network has some branches in their transition routes. When the process comes across a branch, it attempts instantiation of motion schemata with each branch, and it selects the branch at which instantiation is successful. Every motion schema knows the valid context in which to work. Consequently the appropriate route for the current context is chosen automatically. This is the second characteristic of the action network.

Because of these two structural characteristics, the action network can smoothly accept the operator's mental image.

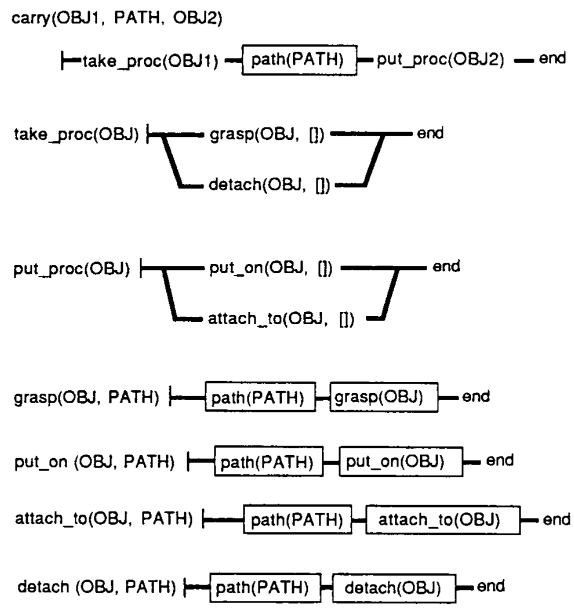

Fig. 4. Example of the action network. The action "carry" consists of sub-actions; take the object OBJ1, go through the path PATH and put onto the object OBJ2. The sub-action "take_proc" is translated as either "grasp" or "detach" according to the context. In this way, "carry" is translated to a list of the motion schemata depending on the context. Motion schemata are marked with squares in this figure.

6. Work Space Model

Geometric data of the real world is captured through the vision system. Although the geometric data containing placement of each object is provided by a CAD

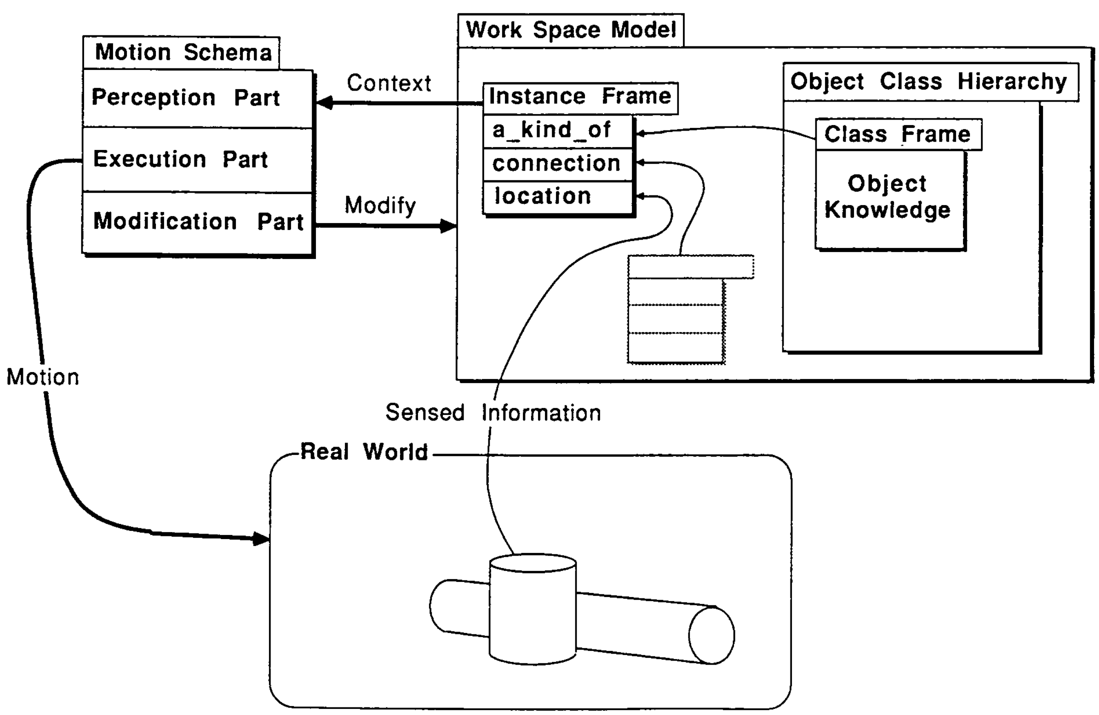

Fig. 5. Structure of the work space model and its interactions between motion schema and the real world.

system, actual placement should be acquired from the real world. But the context of the work space requires additional intelligence. To execute an action, not only geometric data but also knowledge of objects is required. This knowledge includes such factors as grasp point or grasp width of the objects which appear in the motion schema slots. This is the first part of the intelligence. On the other hand, geometric data contained in the system does not always have to be complete and exact. Accurate placement values should be measured when they are needed according to the motion. The important element is that the system always knows the strategy for obtaining exact geometric data from the real world. This is the other part of the intelligence. It is intended that the work space model will have both of these intelligence factors and provide the work space context.

The work space model and its interaction between the motion schemata and the real world is shown in Fig.5. While the information in the real world is only a collection of data, information in the work space model is extracted selectively from the real world and labeled with symbols. The work space model is a kind of cognitive map which describes the real world in the selected domain and the depth of understanding necessary for the expected task. Symbols in the work space model are instance frames corresponding to the objects in the real world. The motion shema is bound to this instance frame, and obtains the slot values from it.

The instance frame has three slots; a_kind_of slot, connection slot and location slot.

The a_kind_of slot indicates the class of the object. The work space model has an object class hierarchy. It contains class frames which describe the knowledge regarding each object class. The instance frame accesses the object knowledge from the class frame by the a_kind_of slot.

The connection slot indicates the connection relationship of the object. It contains an object name and connection type. The object name indicates the frame object it supports. The connection type indicates the tightness of the connection. We prepared three connection types; "on" means that the object is placed on a table or floor, etc., "attached" means that the object is attached but can be easily released, "fixed" means that the object is firmly fixed by bolts or similar means.

The location slot contains the placement data of the object. The value of this slot is updated by the vision system when required. The modification part of the motion schema changes this location slot and connection slot values according to the motion side-effects.

In our situation, the placement data acquisition process is limited to placement error detection. That is, initially approximate values are given by a CAD system or an interactive mapping system [4][5]. Gross changes of object placement are caused by actions, but they are traced in the work space model by the modification of the motion schema. Therefore only small placement errors should be detected by the vision system using the current placement data of the instance frame as an estimated value [3]. When an unexpected gross error is detected, there is a potential for some trouble and the system can execute countermeasure.

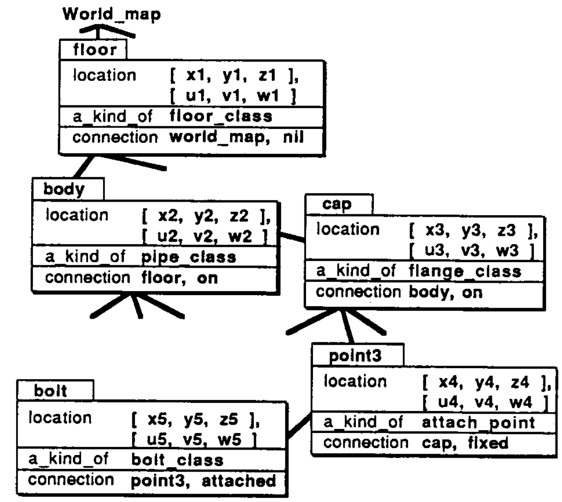

Fig. 6. Example of the work space model. The body is placed on the floor. The cap is placed on the body. The bolt is attached to point3 of the cap.

All the detected placement data and work space context changes are reflected in the work space model. Every independent motion schema works to obtain the context from the work space model. As a result, each motion schema can work consistently . In this sense, the work space model works as a kind of blackboard system. An example of the work space model is shown in Fig.6.

7. Implementation and Experiments

We implemented a system closed in the computer software, as a prototype system of the image based robot operation, and conducted some experiments. Because of our computer hardware restrictions, the robot actuator driving process and the vision process are not combined with this system yet. But using the motion schema, the action network and the work space model, reasonable robot motions are generated and displayed on the system screen as animations. The motion schema, the action network and the work space model are written in Prolog language.

The square frame displayed near the robot gripper is a cursor as shown in Fig.7. The operator specifies the objects to be handled by moving this cursor. He makes an action message specifying objects and words such as "carry" or "grasp." Following the aforementioned process, the system generates an instant motion trajectory. On the lower right window in Fig.7, generated trajectory is displayed as a kind of robot language. In this experiment,

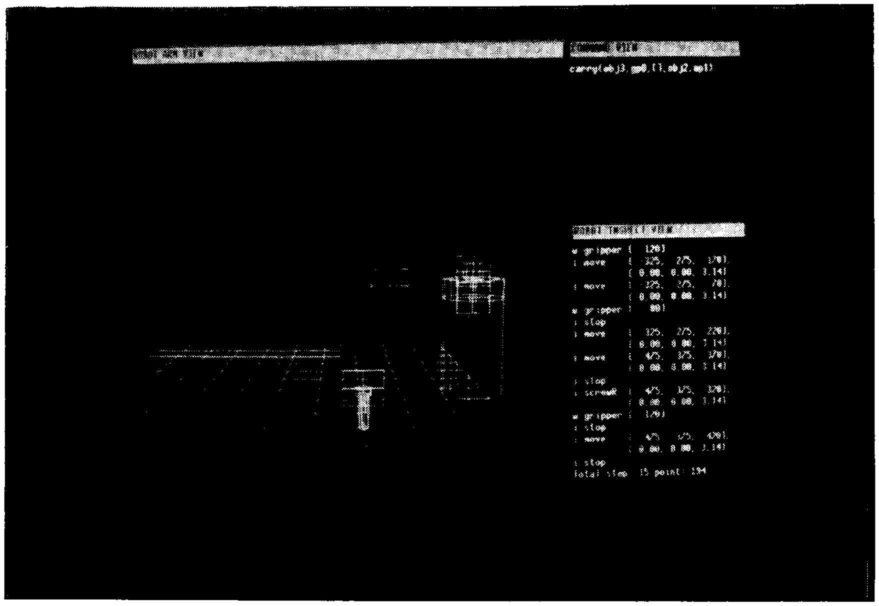

Fig. 7. System Screen Display.

three objects; a body, a cap and a bolt are placed on the floor. The operator instructed the robot to put the cap on the body and fix it with the bolt.

8. Conclusion

A knowledge representation method for an image based robot operation was proposed. Motion concepts which can not be defined by instant motions are consistently represented by the motion schema. In this framework, essential functions of the motion are formalized in the three aspects; perception of the work space context, execution of the instance motion and modification of the work space context. The context perception ability in particular plays an important part in the self-controllability of each motion schema.

As the result of symbolizing motion concepts in motion schemata, actions are described in a generalized style by the action network similar to generative grammar of natural language. It is interesting that an entirely different intelligence; action and language can be represented in the same syntactical structure. We can surmise that this syntactical structure of knowledge is one of the basic features of human intelligence.

The work space model functions as a map for the geometric data. But its more significant role is as a map for the context. Every motion schema is directed by the context, and the context is directed by the internal knowledge of the objects and external cognition; the placement detection. In addition, the external cognition is directed by the real world. Therefore the whole system can work based on the real world situation.

Through implementation and experiments, certain aspects of the proposed method were verified. Total evaluation, however, must be accomplished after completing the combination of the system, i.e., the vision process and the robot mechanism. We anticipate obtaining more conceptual knowledge from this practical experiment.

References

[1] T. Hamada and K. Kamejima, "Image Based Operation - Robot Arm Trajectory Generation from Summarized Motion Planning -," *Proc. 3rd SICE Symp. on Human Interface*, Oct. 1987, pp. 149-152.

[2] U. Neisser, *Cognition and Reality - Principles and Implications of Cognitive Psychology*. W. H. Freeman and Company, 1976.

[3] K. Kamejima, Y. C. Ogawa and Y. Nakano, "Perception - Control Architecture in Image Processing for Mobile Robot Navigation System," *Proc. SICE-IEEE IECON'84*, Oct. 1984, pp. 52-57.

[4] K. Kamejima, Y. C. Ogawa and Y. Nakano, "Image Structure Detection with Application to Mobile Robot Navigation," *Proc. IEEE IECON'86*, Sep. 1986, pp. 713-718.

[5] Y. C. Ogawa, K. Kamejima and Y. Nakano, "Syntactic image Analysis for Environment Understanding," *Proc. IEEE Int. Workshop on Industrial Application of Machine Vision and Machine Intelligence*, Feb. 1987, pp. 266-271.