SPEECH CONTROL FOR MOBILE ROBOTIC SYSTEMS

Arkady S. YUSCHENKO, Dmitry N. MOROZOV, Andrey A. ZHONIN

Bauman Moscow State Technical University, 105005, Moscow, 2-nd Baumanskaya str., 5, Russia

robot@bmstu.ru, morozovd@mail.ru, neurofish@yandex.ru

Abstract: The experience and intelligence of human are necessary to fulfill the hazardous and responsible operations by mobile robot in undetermined environment. To make the control process more effective and simple for human the speech control may be used. The operator's interface in this case may be created using the linguistic variables both for commands formalization and for information presentation. The speech controlled robot has to be an autonomous intelligent system capable to re-cognize the current situation and to adopt its behavior to real environment. To adopt the artificial intelligence to the human impression and reasoning the fuzzy logic principles may be used to create the knowledge base of a speech controlled robot. The simple manipulation and locomotion operations may be presented in form of fuzzy production rules. For complicated modes of behavior the procedure of fuzzy AI – planning have been proposed. The procedure of robot learning on the base of fuzzy neural networks has been developed .for the situations when human-operator can not formalize the fuzzy rules of robot behavior beforehand.

1. INTRODUCTION

Mobile robots are widely applied for complicated operations in undetermined environment. Among such operations are mine disarming, fire fighting, rescue operations, medical service, etc. Robotic systems are normally equipped with manipulators, different sensors, including vision systems. Such systems are generally controlled by human operator whose experience and intelligence is necessary to fulfill the hazardous and responsible operations. The control of mobile robots is often realized now as a remote control by human-operator using a kind of joystick to guide the motion of a manipulator or a chassis. This mode of control is difficult for operator working in time and information deficit and often inefficient creating the risk of error due to the "human factor". Thus a task emerged to arrange the control process so that the operator would have to specify only the aim of the operation. The robot is supposed to assess the environment and make decisions that are necessary to ensure the aims posed by the human operator are achieved. Such systems have been traditionally defined as Intelligent Robotic Systems (IRS). It is rational to arrange the IRS control based on speech, which is expected to take the shape of bi-lateral dialogue between the robot and the operator using a problem-oriented language similar to a natural one.

The speech control, in turn, induces a whole range of artificial intelligence control problems including environment scenes and operator's commands recognition, motion planning, knowledge accumulation and IRS training. Theses tasks are suggested to be solved using the same approach based on linguistic variables and fuzzy logic application. Some of the tasks are discussed below.

2. ENVIRONMENT REPRESENTATION

The environment representation in a human-controlled IRS is based on the corresponding representation in human mind. The information necessary for control in a given situation can be expressed using the means of a natural language (NL) and then translated to a formal language of the relevant semiotic model that is used to control the robot. One of the peculiarities of such representation is its structure, i.e. the representation of the environment as a set of objects bearing particular names and linked with particular relations. D.A.Pospelov names 11 types of such relations (including spatial, temporal, quantitative, causal, and others) (Pospelov, 1986). In most cases of mobile robots control, the environment description in IRS based on fuzzy representations includes the description of objects in the environment as well as spatial and temporal relations between them. The human is known to assess these relations using psycho-physiological scales, defined by objective properties of the corresponding receptors in his body. Therefore the most adequate means to describe the spatial-temporal relations is the apparatus of linguistic variables which uses the same scales.

To describe the current scene, extensional and intentional relations are employed. The former are represented by the relations that describe the location and orientation of objects. For example *a1 is far, to the right, to the fore and a little above a2.* The latter include the relations like *R_1 – to be adjacent to; R_2 – to be inside of; R_3 – to be outside of; R_4 – to be in the centre of; R_5 – to be on the same line as; R_6 – to be on the same plane as; R_7 – to have zero projection on, R_8 – to be on the surface of.* Two unary relations are also proposed in (Pospelov, 1986) – R_{00} – *to be horizontal* and R_{01} – *to be vertical*, as well as 28 elementary spatial binary relations.

The set of specified objects in the current scene, the relations between them, and transformation rules constitute a formal language for scene representation, that is similar to a natural language. Scene description in this language allows for a formal semiotic representation that uses the spatial-temporal relations logic. So, a complex relation *a_1 is on the surface S far and to the right* can be written as *$(a_1\ R_8\ S)\&(a_0\ d_5\ f_7\ a_1)$,* where a_0 – is the observer, with respect to whom the distance and orientation relations are formulated.

Since the environment is ever-changing due the motion of the observed objects as well as to the motion of the robot itself, the scene description changes in time respectively. This circumstance requires that we take into account not only spatial but also temporal relations in the external world, such as *to be simultaneous with, to be prior to, to follow* etc.

3. OPERATIONS DESCRIPTION

External world description allows to pass on to the description of robot's operations within it (Yuschenko, 2002). We assume that complex operations performed by the robot can be represented as a sequence of relatively few typical consistent operations. These are define in advance and are stored in the IRS knowledge base as frames of typical operations. A frame of this kind contains linguistic variables based description of the aims of an operation, the initial stage scene, and the preconditions for the feasibility of the operation. The latter may depend on the specific situation, the capabilities of the robot in question, and the properties of the object of the operation. Thus, the structure of a typical operation frame is as follows: *<operation name> <operation object> < initial situation (modifier of place)> <target situation> <operation feasibility conditions (preconditions)>.<additional details>***.** For example: *<move> <object A> <object A on B> <object A on C> <object A is free>* **<***install object A shock-free***>**. While performing technological operations this frame should sometimes have an extra slot <operation performance method (modifier of manner)>.

Preconditions are one peculiarity of the discussed operations description approach. Generally, all preconditions can belong to one of at least three types: a) situational, e.g. the condition *object A is free* means that *there are no other objects on object A*; b) preconditions stipulated by the robot's capabilities: *the robot is equipped with the gripper suitable for type and size of the object*; and c) preconditions connected with the peculiarities of the object: *the object is a rigid body and can withstand the force developed by the gripper without any damage*.

The description of typical operations expands the situation description language mentioned above. The operator can control a robotic system directly by giving the names and aims of the typical operations in the problem-oriented language, e.g., *‹move object A to plane C› ‹insert shaft A into orifice O›*. Preconditions description may not always be complete in the sense that some of them may not be defined. For example, it may not be known whether there is free space on plane C, on which object A is to be put. Then a query to the cognitive operations base is formed, and an operation is selected for examining plane C that is supposed to provide for filling in the empty slot. The system can also formulate address queries to the operator, if cognitive actions yield no results or uncertainty persists.

Taking into account the similarity between the proposed language for IRS operations description and the situational control language as formulated by D. A. Pospelov, we shall keep to this term bearing in mind the above mentioned peculiarities of the language for IRS.

4. COMPLEX OPERATIONS PLANNING

We shall use the term complex for the operations that can be represented by a sequence of consistent typical operations that result in achieving the aim. Consistency of operations means that the situation achieved as a result of n-th operation meets the preconditions for the (n+1)-th operation. If after the actual completion of n-th operation the consistency is not achieved, the planning process is repeated, with the current situation being assumed as the initial. A distinctive feature of planning procedure in robotics, as compared to numerous methods of artificial intelligence planning, is the possibility of continuous comparison of the real situation observations and the conditions defined during the planning stage. The comparison can be performed as that of linguistic descriptions of the observed and expected (existing only as a statement) situations. The emerging conflict induces a plan of actions aimed at solving it and hence realization of the desired situation. Thus the aim and the name of each separate typical situation gene-rated by the system based on the comparison of real and expected situations, rather than specified by the operator.

The conflict resolution approach is rather similar to human cognitive activity while planning actions, which is also based on comparing the operative image of the situation and the target image. The conflict resolution principle application requires a further extension of IRS control language. Besides the "vocabulary" of typical operations we now need a "vocabulary" for situational conflicts resolution by means of performing typical operations. If spatial relations are intentional then each type of conflict induces its own typical operation to resolve it.

For example: if the aim is: *$(a_1\ R_8\ S)$,* i.e. object a_1 is on the surface S, while in fact *$(a_1\ \neg R_8\ S)$,* then the conflict induces a typical operation *move a_1 to S* . If the aim is defined as *$(a_1\ R\ _2\ C)$,* i.e. *shaft a_1 is inside orifice C*, while observation results show *$(a_1\ \neg R\ _2\ C)$,* then a typical operation is induced: *insert a_1 into C.* If the condition *a1 is free* is necessary for further operations, while in fact we have: *$(a_2\ R_8\ a_1)$,* i.e. *a2 is on a1*, then a typical operation *remove a_2 from a_1* is induced. One can easily proceed with this list of action that resolve intentional type conflicts.

If the relations are extensional there is no need for a special vocabulary for matching the situation with the required typical operation. Conflict can be resolved by performing a typical operation aimed at the relation specified as its precondition. If a mobile robot R is expected to in position *$(R\ d_1\ \ f_1\ \ N)$* with respect to observer *N*, while in fact a different conditions holds true: *$(R\ d_2,\ f_2\ \ N)$,* then the required operation will be defined in the form of: *move robot R from position (R d_2 f_2 N) to position (R d_1 f_1 N).*

While planning complex operations there emerges a multi-step procedure of conflicts resolution. At first, the target and the actual situations are compared. If they do not coincide, the conflicts are defined and the actions are devised to resolve the conflicts. Then the preconditions of the resolving actions are checked, as they can also be in conflict the actual situation. They generate new actions and so on, until at least one resolving action meets the necessary conditions. Then this operation is performed (so far on the planning level), and a new situation appears, which is analyzed in a similar way and so on. This procedure can be represented as directed graph, with its root being the target situation (Yuschenko, 2005).

A disadvantage of the existing approach is that the operator has to define the rules for different situations beforehand, hence the situations should also be known in advance. If the operator fails to formulate the IRS operating rules, then the system can be taught instruction. In this case the operator guides the robot through typical situations after which the information is processed in, e.g. teachable fuzzy (hybrid) neural networks (Vechkanov et al., 2002).

5. SPEECH INTERFACE

Speech interface is the main method to transfer the control data to the IRS. In consists of recognition and linguistic blocks. The recognition block is a device for transforming speech signals as well as interpreting them as separate words or phrases. The linguistic block performs the interpretation of statements into situational control language, as well as the representation of these statements in a semiotic form.

At present there are two most widely used methods of speech recognition : Dynamic Time Warping (DTW), or template matching, and hypothesis probability estimation using Hidden Markov Models (HMM). The template matching method can hardly be regarded a continuous speech recognition method. Moreover, it is speaker depen-dent and requires periodical templates refreshment. For continuous speech recognition one can employ template phrases construction, using the information on the grammar of the IRS problem-oriented language. It is possible to increase the number of operators whose commands the system can efficiently recognize, by means of the so-called method of multiple templates.

The HMM method using hypothesis probability estimation with Viterbi algorithm (beam-search) allows to recognize continuous speech almost independent of the speaker. However this method requires a high quality and expensive teaching speech database. Moreover, the hypothesis probability estimation method implies that the a-priori probability distribution of different hypotheses is known in advance, at the same time ignoring the possible similarity of speech messages. In other words, the HMM method is incapable of detecting and using the distinctive features of words or phrases, in contrast with the template matching method.

Operators statements for IRS control can be formulated in the robot situational control language mentioned above. The linguistic analyzer performs the syntactic and semantic decomposition of the statement which is supposed to result in filling the slots of the frame that describes operations.

When passing from speech sound signal recognition to the inner representation of the operator's sentence, the sequence of words-members of the sentence undergo a formalization procedure. Each sentence – except degenerate commands like "stop" – is presented in the form of typed predicative structure. While describing a sentence, that is in fact a command to perform a certain operation, the corresponding verb plays the key role and is described by the higher-order frame. The slots of the frame are filled in with relevant subordinate parts of the command-sentence. Relying on L. Tesniere's *verb-centric theory* we can introduce the obligatory and arbitrary valences for each of the verbs that describe the IRS' operations. Each slot of the higher-order frame can also be an encapsulated frame, which is the case, e.g. with operation object description. This relieves the operator from the necessity to include into the command-sentence all available information on the object in question. Providing only one of the identification tags allows to assign to the object all available data (size, position, e t. c.) on the semantic level of the speech interface without human interference. The linguistic recognition stage output is a set of encapsulated frames that can be uniquely interpreted over the further stages of command-sentence completion.

It was shown above that the sentences represented by linguistic frames can be expressed in the inner semiotic language as a sequence of symbols. Operator's command that arrives through the speech recognition block is in turn a sequence of symbols as well. Thus, the interaction between recognition and linguistic is reduced to transforming one sequence of symbols into another, based on an expert-built grammar. At the same the linguistic analyzer can be represented as finite-state automaton.

The speech recognition block generates a stream of word hypotheses constituting the operator's sentence. For each word a hypothesis is selected that has the higher probability value. The linguistic block after recognizing each separate word is to define a manifold of acceptable ending of the phrase, with each variant of the complete sentence being assigned the corresponding probability. Using this information the recognition block selects the most plausible hypothesis for the next word (or phrase). The advantage of using the DTW method in this case is in the fact that the reduction of number of the hypotheses to be recognized allows us to use the computational adjustment procedure that would choose the most relevant hypothesis of the remaining few.

Note that the recognition block can in some cases result in recognition failure instead of a hypothesis, as is the case when the noise level is high. In this case the IRS requests the operator for the missing information. The operator's answer can in turn pose new questions which brings about the requirement to fit the linguistic block with a separate dialogue-planning module.

6. CONCLUSION

The preliminary research has shown that the implementation of speech control for a robotic system by way of formulating separate commands is inefficient. It is necessary to develop a speech interface meant focused on the use of problem-oriented language similar in its structure to the situational control language. This allows a substantial simplification of the task of robot control, as it no longer requires any special skills from the operator. There are ho-wever a number of tasks in this field that are yet to be solved. In particular, the application of speech interface for teaching the robot, rather than merely controlling it, when the rules of behavior cannot be formalized in advance, is seen as very important. We also attribute crucial importance to the psychological aspects of interaction between a human and an “intelligent” system, connected with “mutual” ideas about the situations and reasonable beha-vior.

REFERENCES

1. **Pospelov D. A.** (1986), *Situational Control: theory and practice*, Nauka – Phys.matt.lith., Moscow.
2. **Vechkanov V. V., Kiselev D. V., Yuschenko A. S.** (2005), Adaptive system for mobile robot fuzzy control, *Mekhatronika*, Vol. 1, 20-26.
3. **Yuschenko A. S.** (2002), Robot distance control using fuzzy concepts, Iskusstvennyi intellect, *Vol. 4, NAS Ukraine*, 388-396.
4. **Yuschenko A. S.** (2005), Intelligent planning in robot operation, *Mekhatronica*, Vol. 3, 5-18.