Synchronized Multimodal Expression Generation using Editing Toolkit for a Human-friendly Robot

Woo Hyun Kim, Jeong Woo Park, Won Hyong Lee, Won Hwa Kim, Myung Jin Chung, Member, IEEE

Abstract—Attempts to put robots to practical use have been increased, as robots become more human-friendly. In the human-robot interaction field, main issues are how variously the robot can express its emotion and how much the expression is socially acceptable. This paper proposed the editing toolkit which allows us to simulate a 3D model robot in order to express robot’s emotions and intentions for human–robot interaction and robot services. Using the editing toolkit, we have generated multimodal expressions and formulated the method to combine a few of the primitive expressions. The robot, which we used for simulation, has three modalities such as a facial expression, neck motion, and a gesture with two arms. The expressions of each modality were used for generating multimodal expressions, synchronizing with the time information obtained from a professional actor. Consequently, for three emotions and thirteen intentions of the robot, we have generated primitive expression database set and synchronized multimodal expression using editing toolkit.

I. INTRODUCTION

RECENTLY, as the entertainment robot and the service robot are catching popularity gradually, many researchers concern about development of the robot which is able to generate and express its emotion like a human. When a robot interacts with human, if the behavior of the robot seems to be lively and varied, the robot can promote intimacy with human as well as communicate with human more accurately and clearly.

When a robot expresses its emotion and intention, multimodal expressions are great help to interact with human. If a robot expresses facial expressions with neck motions, gestures, sound or LED, the robot can continue to give expression without any sense of boredom. However, for satisfactory multimodal expressions, determining how the various types of modalities are synchronized has to be considered in multimodal expression, because the expression of each modality should start displaying differently and combine more than one primitive expression of each modality together.

The editing toolkit, therefore, is necessary to modify primitive expressions of each modality, generate synchronized multimodal expressions and simulate a robot. In addition, the editing toolkit is allow to manage position and orientation of each joint, the velocity change and other features of the trajectory as well as save and load that information in database form.

Numerous and consistent attempt to develop expression robots have been made, such as implementing animal, machine or humanoid forms on robots. In previous works, using an editing toolkit, we are allowed to handle motors manually for creating new motions. So that, the expression can be modified in detail but cannot be generated automatically and diversely depending on robot’s emotion and intention.

WE-4RII: Waseda Eyes no.4 Refined II has been developed by Miwa and Takanishi [1]. WE-4RII is able to display multimodal expressions corresponding to its emotional state and its personality. However, multimodal expression seemed to be awkward since it has been simply interpolated in order to synchronize.

URBI: Universal Robotic Body Interface has been developed by J.C. Baillie [2,3]. He proposed URBI as the standard for robot control and interfacing which would be both powerful and simple to implement. URBI has been applied to AIBO ERS7 and analyzed the performance of this implementation. URBI is originally an interpreted and scripted low level command language which allows user to read and set motors and sensors directly. For the better performance of robot, it provides high level functionality and state machines. On the other hand, it brings about an increase in complexity as a number of behavior states and a number of conditional sentences increase.

In this paper, the new type of the editing toolkit has been proposed for generating synchronized multimodal expressions of the human-friendly robot. The robot, which we used for simulation, is the mascot-type robot which has three modalities such as facial expressions, neck motion and gestures with two arms. For each modality, we have generated primitive expression database which are used to generate various multimodal expressions through the combination and modification of features in database.

The remainder of this paper is arranged as follows. In the Section 2, main ideas in the generation of the multimodal expression with the proposed editing toolkit and a process for combining primitive expression database are described. In the Section 3, the formation of primitive expression database is defined. Also, in this section, how to use the proposed editing toolkit is described. In the Section 4, finally, the simulation results are shown through the 3D model based robot.
II. GENERATION OF THE MULTIMODAL EXPRESSION

In order to generate various multimodal expressions, determining the position and the orientation of all joints every time might be routine work. In addition, that way the robot cannot generate by itself and the expression of each modality cannot be synchronized without any control by users. Hence the new type of the editing toolkit should be able to combine and modify primitive expression database which have been already created by users. This idea make the multimodal expression more variety and lively. Overall procedure of proposed method is as follows.

A. Gathering information about human expression

Based on information about human expressions, robot expressions can be determined reflecting hardware design of the robot. For the reliability of expressions, professional actor expression data have been obtained using the motion-capture sensor and bibliography of the behavior science. From the professional action captured, we can extract facial expressions, neck motions and arm motions as well as time sequence of this action. When the actor makes motions to express chosen emotions and intentions, we cannot be aware of a delicate difference of the facial expression, so that we used previous works with respect to facial expressions corresponding to emotions [4]. Robot facial expressions were determined by the information through questionnaires of about thirty people in various fields.

![Fig. 3. Facial Expression corresponding to emotions (From the top/left, angry, surprise, happy, disgust, sad, fear)](image)

B. Generating primitive expression database

To produce diverse expressions automatically, it is necessary that the primitive expression is separated from the multimodal expression that we obtain from the professional actor. If primitive expression is not defined, we will have to generate multimodal expression in all cases. Each primitive expression doesn’t have any meanings of robot’s emotion and intention. For example, there is “Raising right hand”, “Extending both arms”, “Turning head on the left side” or “looking up head”. Thus we just mix primitive expressions to generate multimodal expressions. In addition, we just change features of the primitive expression included in multimodal expressions so that we can produce diverse expressions and furthermore, robots can express their emotions and intentions.

![Fig. 4. Primitive expression example](image)
be changed. The time information includes starting time, maintenance time for expression, and required time for reaching the target position of motors.

C. Combining and synchronizing primitive expressions

Generating the multimodal expression is very delicate works because it rarely occurs that people accept the robot’s expression and feel friendly toward the robot. To satisfy human demands, robot’s expression should be various and well-defined with robot’s multimodality. We proposed the effective method to express robot’s emotion and intention, combining and synchronizing primitive expressions. This approach saves time to make robot’s various expressions and allows the robot to generate expressions itself.

III. THE EDITING TOOLKIT

A. Formulation of primitive expression database

We have formulated primitive expression data structure as follows. One primitive expression data include motor trajectory and features. The trajectory has several target positions of each joint as many as robot needs. Features are parameters for diversifying the primitive expression and showing expressions realistically to people, such as each motor speed, starting time, maintenance time for expression, and required time for reaching the target position of motors. Features get primitive database to avoid duplication and reduce user’s works to generate expressions. Primitive expression database is stored in text file form and modified using the editing toolkit.

B. Features of the editing toolkit

We have implemented the editing toolkit in C++. The editing toolkit include 10 slider bars connected with facial expression, 2 slider bars connected with neck motion and 12 slider bars connected with two arms motion. In the editing toolkit, there are also several buttons and edit boxes to save robot’s expressions in database form. Below the slider bars, there is a list box to show the configuration of robot’s expressions.

We can drag the slider to control each joint of the robot and input the information such as motor speed, starting time, maintenance time, and so on. In order to make expressions lively, it is necessary to generate trajectory of the motor rather than one target position. Proposed editing toolkit allows us to make many target positions of each motor within the limits of the possibility of computer memory space as well as it allows us to modify each target position and time information explained the section 2.2 in order to synchronize primitive expressions.

Figure 5 is illustrated how the multimodal expression is generated. Among the pre-defined primitive expression database, we choose the necessary expressions to express certain emotion and then modify features of primitive expressions to synchronize and diversify. It is possible that several expressions are used for one modality, because primitive expression doesn’t imply any robot’s emotions and intentions and multimodal expressions are often constituted of a few repetitive motions.
IV. SIMULATION RESULTS

We have conducted simulations to evaluate about 50 robot’s multimodal expressions. We have classified about 50 robot’s expressions under 13 intentions, idle motions and body language. Because the robot has been designed in order to express their emotion and intention with voice, we defined 13 robot intentions according to means of a sentence that robot say as table 1, and we made one robot expression for one intention. We used a real time 3D dynamic simulation platform for articulated multi-body systems, SimStudio developed by SimLab Co. [5], as following figures.

<table>
<thead>
<tr>
<th>Type</th>
<th>Robot Intentions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Closed Question</td>
<td>Do you want to know weather forecast?</td>
</tr>
<tr>
<td></td>
<td>Open Question</td>
<td>What would you like to drink?</td>
</tr>
<tr>
<td>Asking</td>
<td>Suggestion</td>
<td>Why don’t you play game with me?</td>
</tr>
<tr>
<td>Response</td>
<td>Request</td>
<td>Select a user name.</td>
</tr>
<tr>
<td></td>
<td>Approval</td>
<td>Ok.</td>
</tr>
<tr>
<td></td>
<td>Refusal</td>
<td>No, I can’t do that.</td>
</tr>
<tr>
<td></td>
<td>Positive Response</td>
<td>How wonderful!</td>
</tr>
<tr>
<td></td>
<td>Negative Response</td>
<td>I’m missing you.</td>
</tr>
<tr>
<td>Explanation</td>
<td>Greeting</td>
<td>Hello! I’m a H-robot.</td>
</tr>
<tr>
<td></td>
<td>Opening</td>
<td>Welcome to game world.</td>
</tr>
<tr>
<td></td>
<td>Closing</td>
<td>Good-bye.</td>
</tr>
<tr>
<td></td>
<td>Noticing</td>
<td>Clicking the button goes to the next screen.</td>
</tr>
<tr>
<td></td>
<td>Informing</td>
<td>An interview has been scheduled for you at 7 pm.</td>
</tr>
<tr>
<td>Idle Motion</td>
<td>Waiting</td>
<td>Waiting, Pausing, Habit</td>
</tr>
<tr>
<td>Body Language</td>
<td>Emphasizing the word</td>
<td></td>
</tr>
</tbody>
</table>

Each expression according to the robot intention is able to be changed by robot’s emotions. We have categorized robot emotions into three classes such as positive emotion, neutral emotion, and negative emotion and finally generated total 39 robot expressions according to robot’s emotion and intention.

Figure 6 shows synchronized a multimodal robot expression successively. Because selected primitive motions of each modality have different starting time, we have used the proposed editing toolkit to synchronize.

V. CONCLUSION

This paper proposed the editing toolkit which allows us to simulate a 3D model robot in order to express robot’s emotions and intentions for human-robot interaction and robot services. Using the editing toolkit, we have synchronized generated multimodal expressions and formulated the method to combine a few of the primitive expressions. The expression of each modality was synchronized with the time information obtained from a professional actor. Consequently, for three emotions and thirteen intention of the robot, we have generated primitive expression database set and synchronized multimodal expression using editing toolkit.

There are still some issues for improvement in the proposed method. When combining primitive expressions, we didn’t consider boundaries of primitive expressions. The combination might cause problems if the end position of the former primitive expression differs from the start position of the latter primitive expression. We have two ideas to solve this problem. Firstly, it is the way that primitive expression is modified to avoid a conflict. Secondly, it is the way that the
new primitive expression is inserted between two primitive expressions.

We might be able to generate expressions variously if more diverse features have been defined. Now, we regard trajectory of each motor, motor’s speed, repetition, time information for synchronizing as features of the primitive expressions. If we analyze the trajectory of each motor, we can extract some information to be features such as smoothness, variance and amplitude.

REFERENCES


[5] SimStudio simulation platform : www.simstudio.co.kr


