Title: Phrase-based Gesture Type Inference and Gesture Generation for a robot

1st Yu-Jung Chae· 2nd Hunseob Sin· 3rd ChangHwan Kim, · 4rd Sung-Kee Park
Korea Institute of Science and Technology
h15511@kist.re.kr ·116015@kist.re.kr ·ckim@kist.re.kr ·skee@kist.re.kr

Abstract The generating of robot gestures is essential to interact socially with users in Human-Robot Interaction. This paper introduces a new methodology to generate robot gestures by recognizing phrases and inferring gesture types. The methodology is implemented based on tree models using grammatical information in order to reduce word dependency, in contrast to other research that applies a rule-based system using patterns of words.

Keywords Robot gesture · Gesture generation· Phrase recognition· Gesture type inference

1 Introduction

Gestures enhance comprehension of speech and the ability to recall something from memorie, as they provide an image related to delivering a message and maintain a listener's attention [1]. In addition, robot gestures provide users with familiarity and likability with a robot [2]. In this regard, even though it is important to consider the start and end timing of gestures in order to deliver clear meaning and emphasize the important parts of the spoken dialogue with gestures, these studies are not often done. Furthermore, it is important to consider gesture types, since the outcome of the interaction, such as information recall, gesture effectiveness, and naturalness between a human and a robot is different according to each gesture type [3]. However, the most commonly used method is generating robot gestures based on manually predefined scrips from a user or rule-based method. This paper introduces a new methodology for the generation of co-speech gesture suitable for the spoken dialogue of a robot to deliver clear meaning to a user. Our method segments a sentence into phrases to decide the start and end timing of a gesture, and then infers gesture types for each phrase based on the hierarchical tree model using grammatical information, including Part-Of-Speech (POS), Constituent, and the Stem of a word.

2 Current Gesture Generation Approaches

There are several studies to generate co-speech gestures considering the spoken dialogue of a robot and virtual human. Le et al., proposed a method to generate gestures based on predefined scripts from a user [4]. This method can be applied to make appropriate gestures easily by assigning gesture names in a particular part of a sentence. However, this method needs to manually predefine the gesture names and detailed motions, such as the palm orientation and hand shape for each gesture. In order to automatically generate robot gestures, Ng-Thow-Hing et al., suggested a system using patterns of a word(*e.g.*, between ~ and ~); specifically, the patterns were used to choose gesture types based on predefined rules and the keywords were used to assign a particular gesture [5]. Therefore, it also requires a lot of time to define the rules, and it is impossible to recognize the gesture types when a new sentence, which is not in the rules, is entered. Kim et al., also proposed a method to automatically generate gestures based on patterns of the POS and stem of the word [6]. The patterns of the POS are used to select important words to be expressed with gestures. Gestures are then assigned according to each stem of the word in the selected pattern. However, the patterns of the POS are restricted to constructing a single word. As a result, selecting word groups, which have semantic relations between words, are not considered by Kim et at,. Furthermore, it is not easy to express a natural gesture during the play time of a word due to lack of time.

3 Definition of robot gesture types

We redefine gesture types for a robot based on the theory of human gesture types by McNeill (Table 1) [7]. Figure 1 shows snapshots of each gesture type where we collect human gestures based on our definition by using motion capture cameras. These human gestures are automatically converted to robot gestures by the joint matching method in order to generate natural robot gestures.

Table 1. DEFINITION OF GESTURE TYPES FOR A ROBOT

Gesture Type	Content	Characteristic
Iconic	Describing images of objects or static states	Small &Slow gestures
Metaphoric	Expression of actions or dynamic states	Large & Fast gestures
Deictic	Indicating positions or directions of an object	Pointing gestures
Beat	Strongly highlighting important parts	Repetitive gestures

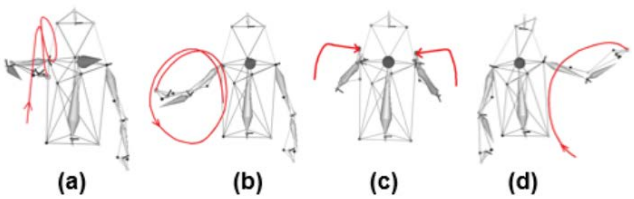

Fig. 1. Snapshots of each gesture type: (a) Metaphoric (e.g. eat) (b) Beat (e.g. too) (c) Iconic (e.g. small) (d) Deictic (e.g. there)

4 Gesture Generations for a robot

The proposed method automatically converts a text to the synchronized synthetic speech and gestures (Figure 2). A sentence is divided into phrases, where synthetic speech with inserted pauses between phrases is generated to make

clear speech and secure more play time for the robot gesture. As a result, an important phrase or phrases are selected considering the relation between the user's personality and robot gesture types. Finally, the gesture is assigned using the stem of a word in the selected phrase.

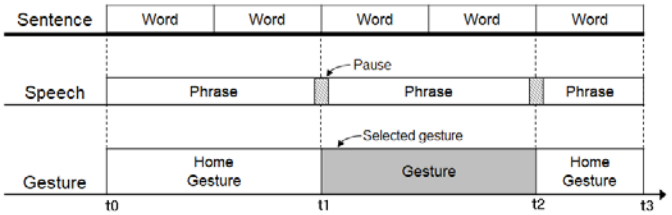

Fig. 2. Concept of our methodology

In particular, the morphological analyzer extracts grammatical information from a sentence that has been inputted (Figure 3). By using this information, the hierarchical tree model segments the sentence into phrases and infers gesture types for each phrase. The speech synthesizer makes a synthetic voice and computes play time for each phrase. The gesture generator selects the important phrases that are expressed by gestures considering the relation of a user's personality with robot gesture types and the gesture name using the designed weight function based on the research [8]. The gesture generator then makes the gesture profile for each phrase, which comes from the primitive gesture DB, and at the same time considers the length of play time of the selected phrase. Finally, the synchronizer simultaneously plays the speech file given by the speech synthesizer and gesture profile from the gesture generator.

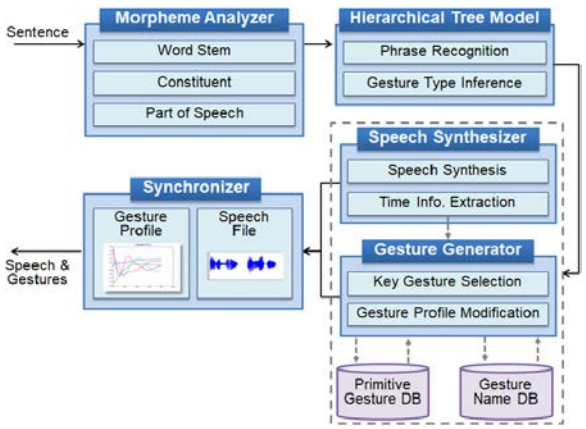

Fig. 3. Architecture of our method

To implement the hierarchical tree model, we apply the random forest models, since this method is able to reduce variance and make general rules. In the first layer of the hierarchical tree model, the feature vector for the phrase recognition $\boldsymbol{\Theta}_P = (\boldsymbol{\Lambda}, \boldsymbol{\Pi}, \#\mathbf{POS}; \#\mathbf{Constituent})$ is defined, where $\boldsymbol{\Lambda}$ means a sequence vector of the POS, where $\boldsymbol{\Pi}$ indicates a sequence vector of the constituent, where #**POS** is the number of the POS, and where #**Constituent** is the number of the constituent. A set of class is defined by $\boldsymbol{\Psi}_P = \{Phrase, None\}$. In the second layer of the hierarchical tree model, the feature vector for the gesture type inference $\boldsymbol{\Theta}_T = (\boldsymbol{\Lambda}, \boldsymbol{\Pi}, \#\mathbf{POS}; \#\mathbf{Constituent})$ equals to $\boldsymbol{\Theta}_P$ only if $\boldsymbol{\Psi}_P$ is the *Phrase*, and a set of class is assigned by $\boldsymbol{\Psi}_T = \{\mathbf{Iconic}, \mathbf{Metaphoric}, \mathbf{Deictic}, \mathbf{Beat}\}$.

5. Evaluation

5.1 Comparison of learning methods

To compare the accuracy of the hierarchical tree model with other methods, we collected 857 sentences that contain complex and simple sentences from Korean elementary school text books. We applied the 10-fold cross validation, and then performance was computed by the accuracy through f-measure and *Receiver Operating Characteristic* (ROC) area as shown in Table 2.

Table 2 COMPARISON OF LEARNING MODELS

	Phrase		Gesture Types	
	Acuracy	ROC Area	Acuracy	ROC Area
NaiveByes	0.744	0.831	0.783	0.912
SVM	0.730	0.705	0.761	0.754
DecisionTree	0.818	0.874	0.820	0.887
RandomForest	0.830	0.905	0.835	0.934

5.2 User evaluations

We assessed our method (labeled Robot A) by comparing with another method (labeled Robot B) that randomly generated gestures (Figure 4). For this experiment, we collected 30 participants, and they watched 8 sets, in which each set contained a spoken dialogue and gestures for Robot A and B. Finally, we were able to observe the higher scores for Robot A on average than Robot B. We are able to prove the validation of our method as shown the p-values.

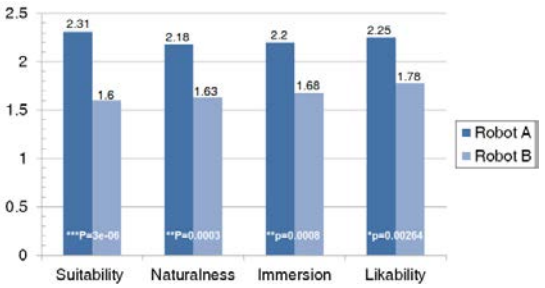

Fig. 4. Results on the questionnaire

Acknowledgement This work was supported by the National Research Council of Science & Technology (NST) grant by the Korea government (MSIP) (No. CRC-15-04-KIST)

References

1. A. B. Hostetter, "When do gestures communicate? a meta-analysis," *Psychological Bulletin*, 237(2), pp. 297-315, 2011.
2. S. Buisine and J.-C. Martin, "The effects of speech–gesture cooperation in animated agents behavior in multimedia presentations," *Interacting with Computers*, vol. 19, no. 4, pp. 484–493, 2007.
3. C.-M. Huang and B. Mutlu, "Modeling and evaluating narrative gestures for humanlike robots," *Robotics: Science and Systems*, pp. 57–64, 2013.
4. Q. A. Le and C. Pelachaud, "Generating co-speech gestures for the humanoid robot nao through bml," *International Gesture Workshop*. Springer, pp. 228–237, 2011.
5. V. Ng-Thow-Hing, P. Luo, and S. Okita, "Synchronized gesture and speech production for humanoid robots," *Intelligent Robots and Systems (IROS)*, pp. 4617–4624, 2010.
6. H.-H. Kim, Y.-S. Ha, Z. Bien, and K.-H. Park, "Gesture encoding and reproduction for human-robot interaction in text-to-gesture systems," *Industrial Robot: An International Journal*, vol. 39, no. 6, pp. 551–563, 2012.
7. D. McNeill, *Gesture and thought*. University of Chicago Press, 2008.
8. Aly, Amir, and Adriana Tapus. "Towards an intelligent system for generation an adapted verbal and nonverbal combined behavior in human-robot interaction,", Autonomous Robot, vol. 40, no. 2, pp. 193-209, 2016.