KSE 2011

Published by the IEEE Computer Society
10662 Los Vaqueros Circle
Los Alamitos, CA 90720

IEEE Computer Society Order Number P4567
ISBN 978-0-7695-4567-7
Library of Congress Number 2011935572
BMS Part Number CFP1103I-PRT

KSE 2011 the third international conference on knowledge and systems engineering

KSE 2011

the third international conference on knowledge and systems engineering

14-17 October 2011 Hanoi, Vietnam

Editors

Tu Bao Ho
R.I. McKay
Xuan Hoai Nguyen
The Duy Bui

Organized by

Hanoi University
University of Engineering and Technology
(Vietnam National University, Hanoi)

Published by: CPS Conference Publishing Services

Proceedings

2011 Third International Conference on Knowledge and Systems Engineering

KSE 2011

14-17 October 2011
Hanoi, Vietnam

Proceedings

2011 Third International Conference on Knowledge and Systems Engineering

KSE 2011

14-17 October 2011
Hanoi, Vietnam

Editors
Tu Bao Ho
R.I. McKay
Xuan Hoai Nguyen
The Duy Bui

Organized by
Hanoi University
University of Engineering and Technology
(Vietnam National University, Hanoi)

CPS Conference Publishing Services

Los Alamitos, California
Washington • Tokyo

IEEE computer society

IEEE Computer Society Order Number P4567
ISBN-13: 978-0-7695-4567-7
BMS Part # CFP1103I-PRT
Library of Congress Number 2011935572

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
http://computer.org/cspress
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
http://shop.ieee.org/store/
customer-service@ieee.org

IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Individual paper REPRINTS may be ordered at: <reprints@computer.org>

Editorial production by Patrick Kellenberger
Cover art production by Mark Bartosik
Printed by Applied Digital Imaging

IEEE Computer Society
Conference Publishing Services **(CPS)**
http://www.computer.org/cps

2011 Third International Conference on Knowledge and Systems Engineering

KSE 2011

Table of Contents

KSE 2011 Full Papers

IWEC 2011 Full Papers

2011 Third International Conference on Knowledge and Systems Engineering

Wizard of oz for designing hand gesture vocabulary in human - robot interaction

Nguyen Thi Thanh Mai [(1)], Tran Thi Thanh Hai [(2)], Nguyen Viet Son [(3)]
MICA Center, HUST - CNRS/UMI 2954 - Grenoble INP
Hanoi University of Science and Technology, 1 Dai Co Viet St., Hanoi, Vietnam
[(1)] Thanh-Mai.Nguyen@mica.edu.vn, [(2)] Thanh-Hai.Tran@mica.edu.vn, [(3)] Viet-Son.Nguyen@mica.edu.vn

***Abstract* - Recently, human - machine interaction (HMI) becomes a hot research topic because of its wide applications, ranging from automatic device control to designing and development of assistant robot or even smart building at sparser scale. One of the most important questions in this research field is finding out an efficient and natural method of HMI. Among several channels of communication, hand gestures have been shown to be an intuitive and efficient mean to express an idea or to control something. In this paper, we propose a framework to study the behavior of Vietnamese in using of hand gesture in communication with robot. This study allows designing a hand gesture vocabulary for human - robot interaction (HRI) applications. This makes our twofold contributions: (1) a general framework of studying and designing an interaction protocol between human and robot; (2) a basic set of hand gestures that can be used in general situation of HRI. This set of hand gestures has been shown to satisfy both criteria: intuitiveness and recognisability.**

Keywords - Human robot interaction, Hand gesture recognition, Hand gestures vocabulary, Wizard of Oz.

I. INTRODUCTION

Robotics is currently undergoing a major change. In the past, robots employed in assembly lines or well structured environments. Nowadays, we can see the presence of robots in many aspects of everyday life for professional as well as personal assistant services.

To assume the communication between human - robot, many researches on HRI have been conducted. An intelligent robot requires a natural interaction with human. The interaction could be performed via several perception channels like vision, speech, touch, etc. Although significant advances have been made in speech-based interface research, these interfaces will be sometimes impractical in both noise and quiet environment.

Gesture is an intuitive and efficient mean of communication between human and human in order to express information or to interact with environment. In HRI, hand gesture can be an ideal way that a human controls or interacts with a robot. Providing the ability of understand hand gestures to robot will improve the ease and efficiency of the interaction.

In order to interact with human through hand gesture, the robot needs to understand hand gestures. The recognition will be performed by learning examples of gestures of interest and then robot can recognize given a new gesture.

For a successful hand gesture based interaction between human and robot, a vocabulary of hand gestures needs to be defined and a gesture based protocol of communication should be understood by both human and robot.

The designing of hand gesture vocabulary needs to satisfy two requirements: (1) the gestures should be intuitive and comfortable to perform by human; (2) the gestures should be recognizable by the system. Few researchers have dealt with this subject in a comprehensive manner considering human (intuitiveness) and technical factors (recognition) jointly. Finding the recognition accuracy of a hand gestures set is not a problem, as many excellent algorithms exist. Measuring intuitiveness is another matter, as this factor is subjective and must be obtained empirically. This seems to be the bottleneck in the design of an optimal hand gestures set.

This paper proposes a Wizard of Oz (WOz) framework [1] for designing such a vocabulary of basic hand gestures for HRI. The use of WOz technique for acquisition of multimodal interaction patterns has been considered as the essential tool for the multimodal interface design. In our work, the study and the designed gestures set, commonly used by Vietnamese in interaction with robot, helps for building applications based HRI by hand gestures. Our main contributions are: study the behaviors of Vietnamese in communicating with robot by hand gesture; define a hand gesture vocabulary that can be used for general purpose, and some evaluations of hand gesture *recognisability*.The paper is organized as follows: In section II, we analyze some sets of hand gestures proposed in the literature. In section III, we propose a framework for designing a hand gestures vocabulary. Section IV represents more detail each step to be performed in order to obtain the results, and some evaluations. Conclusions and future works are discussed in section V.

II. RELATED WORKS ON HAND GESTURE VOCABULARY

Since recent several years, a lot of researches in human - computer interaction based on hand gestures have been conducted [2], [3], [4], [5]. In general, each work has been evaluated on a common hand gesture database that was experimented with another database built by the authors themselves. Some of databases are published for research use, but it is necessary to rebuild database for a specific application. In addition, the methodology for designing and building a hand gesture database has not been mentioned in all related scientific papers.

This section will summarize existing hand gestures databases. There are two approaches for acquisitioning

978-0-7695-4567-7/11 $26.00 © 2011 IEEE
DOI 10.1109/KSE.2011.44

hand gestures: glove-based approach [6] and camera-based approach [7], [8]. Glove-based approach requires the human to wear a specific glove. This approach is very fast and accurate. But the human feels uncomfortable when wearing a sensor based glove. In addition, the glove is very expensive. To overcome this drawback, vision based approach uses camera to capture visually hand gestures. Image (in case of static hand posture) or video of hand gesture (in case of dynamic hand gestures) will be registered for further processing. The camera based approach is more convenient for human - robot interaction in ubiquitous perceptive environment.

In the literature, there are ten public hand gestures databases (static and dynamic hand gestures) that could be used for different applications (e.g. hand sign language [9], robot controls [10]). In this paper, we do not want to do a survey on hand gesture databases in general but we focus only on hand gesture vocabularies for the applications of human - robot interaction.

In [2], six hand gestures have been considered to control robot: pointing, thumbing, relaxed, raised, arched, halt. The hand gestures were defined by the angles between the head, the shoulders and the arms.

In [11], the authors used both static and dynamic gestures to control a trash-collecting robot: stop (moving arm into the right position for about one second), follow (wave-like motion, moving the arm up and down), pointing vertical (move the arm from a position up to a position), pointing pow (starting from a position, pointing to an object on the floor, return to the initial position).

In [9], the authors tested with five gesture types: stop, waving right, waving left, go right, go left. The data are collected from video sequences of five subjects. The subjects are led into a room with constant background and instructed how meaningful the gesture looks. They are further instructed to look at the camera and execute the movement.

In [10], a robot is controlled via five dynamic hand gestures: move forward, move forward then right, move forward then left, move backward then left, move backward then right. These hand gestures are built from one or two hands.

In [12], [13], the authors presented a robot Robotinho who plays a role of tour guide in a museum. Arm and hand gestures are both used for communicating with tourists. The hand gestures that human interacts with robot include: waving (one handed gesture), pointing (parametric, one handed gesture), thisbig (two handed gesture to indicate the size of an object), dunno (two handed gesture to express ignorance). A part from using hand gesture, body and head gesture were also considered.

We found that for each specific application, a hand gesture vocabulary has been proposed by authors. Some gestures are common among applications (e.g. waving), some others have different meaning even the hand movement still remains the same. This requires redefining a gesture set for a new application. In addition, this gesture set, as proposed by researchers, is imposed for human without considering if they do this in a comfortable manner or not.

In human - robot interaction, some communication scenarios remain the same for all applications. For example, before controlling or interacting with robot, human needs to call the robot coming near him/her. When human does not have anything else to command, he/she can make a sign to say goodbye or to end the interaction, etc. Therefore, we think that it should be useful to study and to design a common hand gesture set that could be used for general context.

III. DESIGNING A HAND GESTURE VOCABULARY FOR HRI

A. Framework of designing hand gesture vocabulary

The designing of hand gesture vocabulary needs to satisfy two requirements:

- *Toward human in the interaction*: The gestures should be performed intuitively and comfortably by a human.
- *Toward system (robot):* The gestures should be recognizable by the system.

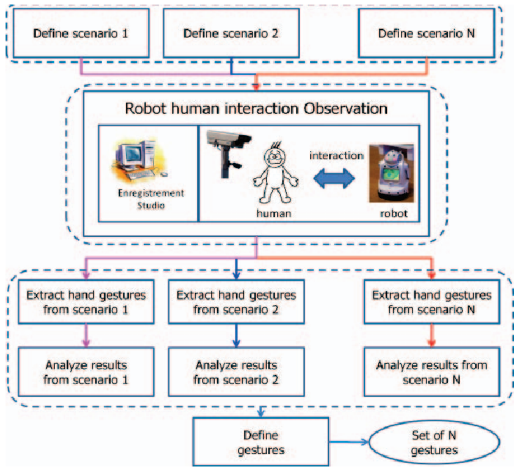


Figure 1. Framework of designing hand gesture vocabulary.

In [14], the authors proposed a method for selecting an optimal hand gesture vocabulary. However, this method is quite analytic and psychological. The authors did not indicate a study case to obtain a vocabulary.

We inspire the idea in this work and propose a framework to design a hand gestures vocabulary in reality (Figure 1). This framework consists of four main blocks: (1) interaction scenarios definition; (2) HRI observation in each scenario by camera; (3) hand gestures extraction and analysis; (4) definition of hand gestures set. In the second block, a set of people will be invited to participate into interaction with the robot without knowing that their interaction is registered (we refer to the Wizard of Oz technique - an efficient way to examine user interaction with robot). This allows obtaining the most natural HRI.

B. Definition of HRI scenarios

In order to study the behaviors of Vietnamese in communication with robot and to build a hand gestures set, we define a series of HRI scenarios in a simulated library context. It needs to be noted that this simulated context is not a special context, so the HRI studied in this context can be used and extended for many other contexts. The scenarios must be basic and simple which allow the subjects play them easily and exactly.

The simulated library is a room of size 3m x 3m in which we equipped some tables, chairs, and bookshelves. All are similar to a reading room in a library so that anyone can feel as in a real library. In this context (the library context), the scenarios are played by two actors (a human and an assistant robot). We focus on some general and basic demands that a human needs and often uses in the library, e.g. a human wants to look for a room, a book/paper/magazine, or he/she wants to know more detail about the library history, the library map, the book/paper/magazine brief … or even wants to know about the robot's services in the library, etc. Figure 2 extracts a scenario frame in which a human is interacting with the assistant robot in the library.

Figure 2. An example of scenario in which a human asks the robot to know more detail about the abstract of a book to which his hand is pointing. The robot answers the human by synthetic voice using Vietnamese speech synthesis system.

To define interaction scenarios, we invent situations and assign roles to a human and a robot. The scenario can start with a human entering into the library, learnt that there is a service robot, he looks around the room to find a robot, then calls the robot coming near to him to ask some services like looking for a book; asking to know more detail about the book; looking for a room; etc. During the interaction with robot, human can do anything (by gesture or voice) to explain his/her demand or his/her attitude to the robot. Once all demands are responded/refused, the human feels (un)happy to pass the time in the library, he/she ends the interaction with robot and goes outside. Figure 2 extracts a frame of a scenario in which a human is interacting with an assistant robot in library.

Although these scenarios are played in the library context with library specific operations, we will study only in the human interacting behaviors with robot in the most five common situations: *call the robot; point to something for a service; agree or disagree with the robot's answer, finish an interaction.* The library context helps the human interacting with the robot in a real situation. To summarize, five interaction scenarios will be considered:

- Call robot to come.
- Point to an object to know more about it.
- Agree with robot.
- Disagree with robot.
- Finish interaction with robot.

C. HRI observation

Once scenarios are defined, we start filming the scene with 3 cameras to assure that all in the scene are visible. A microphone is also used to register voice communication. In order to study the hand gesture set of Vietnamese in HRI, a multimodal corpus (video/audio) was built with twenty-two native Vietnamese people (eleven males and eleven females) with a mean age of 23. There are fourteen right-handers, and eight left-handers. All of them are students.

All people are asked to play two times all the defined scenarios, each at one time. They were asked to express five different demands (call robot, point to something and want to know some more detail information, agree or disagree with robot's answer, finish interaction with robot) by using their voice and hand gestures. All subjects played these scenarios without difficulties, but sometimes they forgot to use hand gestures. It could be explained that Vietnamese are not in the habit of using hand gestures in daily communication.

In order to obtain a natural HRI, we said to the subject that we would like *to test the robot's abilities, i.e. the performance of speech and gesture recognition system embedded on robot while interacting with human*. He/she does not know that robot is controlled by an anonym technician in another room. During interaction with robot, human is asked to do not move a lot such that only hand movement is taken into account.

All twenty-two peoples play two times all the defined scenarios, yielding 66 video files (22 subjects x 3 cameras). All videos files are recorded in the same format avi, sampled at 25 fps with resolution 352 x 280. Depending on the relative position of human w.r.t the camera pose, some hand gestures are visible in one camera field, some others do not. After selecting and editing, we have obtained 850 clips (corresponding to 459 scenarios) that only present one hand gesture per one scenario.

D. Hand gesture analysis

All 850 clips selected from the database are analyzed. The analysis should answer to the following questions:

- Which gestures are used in each scenario?
- How are gestures characterized?

A hand gesture is defined as a sequence of movements of hand postures. In general, a gesture is composed of three phases: preparation; execution; finish [8]. We are interested only in execution phase. We propose to analyze gestures based on hand postures and movement properties during execution phase of hand gesture as movement speed, movement amplitude; performing time of gesture.

- *Movement speed* is defined as hand speed, measured by a distance that a hand moves in a time unit (m/s).
- *Movement amplitude* is defined as the maximum distance between to hand postures centers (Figure 3) when doing a gesture.
- *Performing time:* the total time that a human does a gesture (during execution phase), counting

from the starting point to the ending point.

To obtain the movement parameters of hand gesture during execution phase, we need to:

- Detect boundaries between phases of hand gestures in order to extract only video frames in the time of execution phase from whole video clip.
- Determine 3D hand position at each frame then track the hand movement during the execution phase, in order to compute the amplitude and the speed of hand movement.

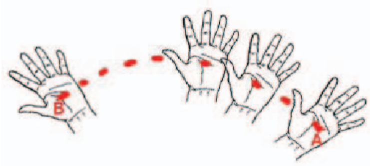


Figure 3. Movement amplitude of a hand gesture is the distance from A (starting point) to B (ending point).

Currently, the boundary detection and hand tracking is done manually. Based on obtained statistic data on movement speed and movement amplitude, the Vietnamese hand gestures in the HRI context could be categorized into three groups:

- *Speed:* Fast (if the movement speed is bigger than 0.5m/s), Average (if its value is between 0.2m/s and 0.5m/s), Slow (if it is smaller than 0.2m/s).
- *Amplitude:* Wide (if the movement amplitude is bigger than 15cm), Average (if it is between 5cm and 15cm), Narrow (if it is smaller than 5cm).

The analyzed results show that:

- We observe two interesting differences between the human - human interaction and the HRI: (1) when interacting with the robot, Vietnamese use hand gesture more often to impress robot than they do human - human interaction; (2) the performing time for one gesture in the case of HRI is longer than the one of human - human interaction, because they need to keep the gestures until obtaining the robot's response. Therefore, in almost scenarios, the amplitude and speed of hand movement are categorized into average group, not narrow and fast group as we expect.
- For each scenario (the same context and the same command), several types of hand gestures are used.

We will now analyze in more detail the percentage that one gesture is used in each scenario. Table 1 represents two types of hand gestures "**Call1** and **Call2**" that Vietnamese use to call robot. Note that the scenario "call robot" is used if human wants to start an interaction with robot or he/she needs a robot's aid. The analyzed results show that 92% Vietnamese uses the **Call1** hand gesture (open, wave, hollow of hand down) to call robot and only the 8% uses the **Call2** (open, wave, hollow of hand up).

In the second scenario (human points to an object to ask more information about it), there are also two different hand gestures that are used in which **Point2** type is used more often (77%) than the first one (23%) (**Point1**) (see Table II).

TABLE I. HAND GESTURES USED TO CALL ROBOT.

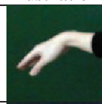
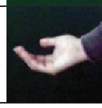
Type	Illustration	Description	Per.
Call1		hand open, wave, hand hollow down	**92%**
Call2		hand open, wave, hand hollow up	8%

TABLE II. HAND GESTURES USED IN POINTING SCENARIO.

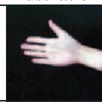
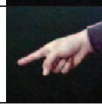
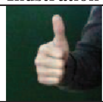
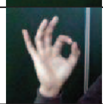
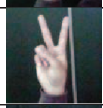
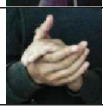
Type	Illustration	Description	Per.
Point1		open, hand point, not change hand shape	23%
Point2		close, forefinger points, not change hand shape	**77%**

Table III represents the hand gestures used to explain an agreement with robot. In this case, four hand gesture types are carried out, in which the **Agree1** and **Agree2** are used more usually, with 61% and 30%, respectively, than the others.

TABLE III. HAND GESTURES USED TO EXPLAIN AN AGREEMENT WITH ROBOT.

Type	Illustration	Description	Per.
Agree1		fingers close, but the thumb up	61%
Agree2		fingers open, but forefinger and thumb close	30%
Agree3		fingers close, but forefinger and middle finger make the victory symbol	4%
Agree4		hand clap	5%

Disagreement and finishing hand gestures are represented in Table IV and Table V.

TABLE IV. HAND GESTURES USED TO EXPLAIN A DISAGREEMENT WITH ROBOT.

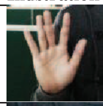
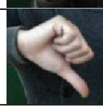
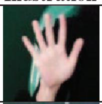
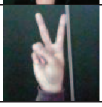
Type	Illustration	Description	Per.
Dis1		Fingers open, hand moves left, then right, then left, not change hand shape	**82%**
Dis2		close, forefinger points down, not change hand shape	18%

TABLE V. HAND GESTURES USED TO FINISH AN INTERACTION WITH ROBOT.

Type	Illustration	Description	Per.
Stop1		Fingers open, hand moves left, then right, then left, not change hand shape	**94%**
Stop2		fingers close, but forefinger and middle finger make the victory symbol	6%

The disagreement scenario is defined in which human refuses or does not agree with the robot's answer. The finishing scenario will be used if human wants to end his/her interaction. One important thing needs to be noted that Vietnamese use two different types of hand gestures in each context, but almost people used one hand gesture (with the fingers open, hand moves left, then right, then left, and shape does not change) to explain two different things (**Dis1** for disagreement scenario, and **Stop1** for finishing the interaction). In the disagreement scenario, the second hand gesture type (**Dis2**) is used quite often (18%), whereas there is only one human who used gesture **Stop2** to finish the interaction with robot (6%).

In order to distinguish these two hand gesture types (**Dis1** and **Stop1**), we carried out some analysis on movement speed, movement amplitude, hand type (right or left), and performing time of gestures.

TABLE VI. THE ANALYZED RESULTS OF DIS1 AND STOP1 IN TWO SCENARIOS: DISAGREEMENT AND FINISH.

Gesture	Speed (%)			Amplitude (%)			Time (s)
	F	*M*	*S*	*W*	*M*	*N*	
Dis1	30	61	9	18	44	38	1.56
Stop1	31	58	11	13	38	49	2.27

Table VI shows that all most Vietnamese carry out the both hand gestures (**Dis1** and **Stop1**) with medium and narrow moving amplitude (44% and 49%, respectively) and medium speed (61% and 58%, respectively). There are 82% and 73% of human who use right hand for **Dis1** and **Stop1** gestures, respectively. This can be explained because there are fourteen right-handers (64%) among twenty-two people in our corpus. So it can be noted that the both two hand gesture types (**Dis1** and **Stop1**) present many similar characteristics, it can make a difficulty to robot for recognizing.

IV. HAND GESTURE VOCALBURARY DEFINITION

As previous speaking, the designing of a hand gesture vocabulary needs to satisfy two following criteria: (1) the *comfortableness* for human; (2) the *recognisability* for system.

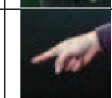
The first criterion is assured by choosing hand gestures that are mostly used by human when interacts with the robot. As analysis above, five following hand gestures will be selected: **Call1** (92%), **Point2** (77%), **Agree1** (61%), **Dis1** (82%), **Stop1** (94%).

We can see that following this criterion, Vietnamese use one hand gesture to express two different things: *disagree with robot* (Dis1) and *end the interaction* (Stop1). In order to assure the recognisability criteria, hand gestures need to be distinct. We found that in "*disagree with the robot*" scenario, two gestures have been used (Dis1, Dis2). We then decide to choose the remaining gesture **Dis2** that is different from the gesture **Stop1** used in "*finish the interaction*", even the number of using **Dis2** takes only 18%. In conclusion, the hand gesture set that assures two criteria is designed as in the Table VII. This set contains the five hand gestures commonly used in five scenarios usually concerned in HRI context.

For the recognisability criteria, we need to build a hand gesture recognition system and test if the system can distinguish all of the proposed hand gestures. The system is composed of two main phases: feature extraction and posture classification, which is built based mainly on the framework presented in the previous work [15]. The feature for hand representation is Haarlike features and classification is carried out using Cascaded Adaboost technique [16]. For learning phase of Cascade Adaboost classifiers, we built a database with the five gestures above collected from twenty Vietnamese. For each gesture, the positive set includes 1200 positive images (image containing the gesture should be recognized) in two different contexts (uniform background and complex background) under neon lighting condition (60 images / person) and the negative set includes 1500 negative images (i.e. images which do not contain any hand gesture). The stage number for training Cascaded Adaboost classifier is 30. For five hand gestures, we need to train five classifiers.

In order to extract features from an image, a multi-scale sub-window is used to scan across many locations on the image. For each sub-window, if captured area has the same features with the one of positive images, respectively, this area will contain the hand gesture.

TABLE VII. PROPOSED HAND GESTURE VOCABULARY (S: MOVEMENT SPEED, A: MOVEMENT AMPLITUDE, T: PERFORMING TIME IN AVERAGE. M: MEDIUM).

Gesture	Illustration	S	A	T
Call		M	M	3s
Agree		M	M	2s
Disagree		M	M	2.5s
Point		M	M	10.5s
Stop		M	M	2s

In order to recognize a hand gesture, an image containing this gesture will be passed to five classifiers. The hand gesture will be classified into one class whose classifier gives the biggest probability.

For hand gesture evaluation, we perform two experiments with the hand gesture recognition system. The first one (dependent subject experiment) is done with two subjects that had participated in the training database construction. The second one is performed with four subjects who did not participate in the construction of the training database (independent subject experiment). For each subject, we take 500 hand gesture positive images and 50 negative ones. The Table VIII represents the recognition result of the both experiments.

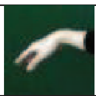
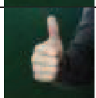
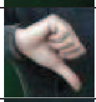
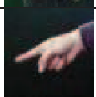
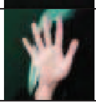
We use recall and precision rate to evaluate the system's performance. In a classification task, the precision and recall for a class are defined as follows:

- Precision = TP/(TP+FP)
- Recall = TP/(TP+FN)

where:

- TP (True Positive) is the item number labeled correctly as belonging to the positive class.
- FP (False Positive) is the item number labeled incorrectly as belonging to the positives class.
- FN (False Negative) is the item number which was not labeled as belonging to the positive class but should have been.

TABLE VIII. THE RECOGNITION RESULTS OF THE BOTH EXPERIMENTS: INDEPENDENT SUBJECT EXPERIMENT AND DEPENDENT SUBJECT EXPERIMENT

Gesture	Illustration	Dependent subject experiment		Independent subject experiment	
		Recall	*Prec ision*	*Recall*	*Prec ision*
Call		89%	93%	93%	96%
Agree		67%	72%	74%	76%
Disagree		98%	96%	93%	88%
Point		92%	95%	89%	87%
Stop		95%	89%	94%	95%
Mean		**88%**	**89%**	**88%**	**88%**

Precision can be seen as a measure of exactness or fidelity, whereas recall is a measure of completeness. In even simpler terms, a high recall means you haven't missed anything but you may have a lot of useless results to sift through (which would imply low precision). High precision means that everything returned was a relevant result, but you might not have found all the relevant items (which would imply low recall).

Table VIII shows the precision and recall results for each hand gesture class. The average value of recall and precision rate is about 88% for the both experiments (dependent and independent subjects). The results show that the proposed gesture set satisfies the *recognisability* criteria. The processing time is about 18fps on a dual core 2.66 MHz, RAM 2GB PC system. It allows the use of the gesture set real-time video applications.

V. CONCLUSIONS

This paper studied the behavior of Vietnamese in using hand gesture to communicate with robot. The study has been carried out through a wizard of oz framework of four steps. The proposed framework is general and could be used for the other studies which aim to study the other interaction methods (e.g. using speech). The results obtained from our study are the five hand gestures set that are used commonly in five scenarios. These hand gestures concern many HRI applications. The designed five hand gestures set satisfies the both criteria: *comfortableness* and *recognisability*. Researchers in the domain can reuse this gestures set without requiring redesigning it.

In the future, we want to build and test some applications of human - robot interaction that will use this hand gesture set in combining with other modality such as speech.

ACKNOWLEDGMENT

This study was done in the framework of the International co-operation project 10/2011/HĐ-NĐT. The authors thank to Van Thai Dong who helps the authors to complete the experiments in this work.

REFERENCES

[1] Nils Dahlbäck, Arne Jönsson, Lars Ahrenberg: Wizard of oz studies - why and how. In Proceeding IUI '93 (Proceedings of the 1st international conference on Intelligent user interfaces ACM New York, USA 1993)

[2] D. Kortenkamp E.H., R. P. Bonasso, and M. Inc: Recognizing and interpreting gestures on a mobile robot. In Proc. of the Thirteenth National conf. on Artificial Intelligence, (pp 915–921). (1996).

[3] A. Eliav T.L., Y. Parmet, H. Stern, J. Wachs, and Y. Edan.: KISS human-robot interfaces. 18th International Conference on Production Research (ICPR). (Salerno, Italy 2005).

[4] Lee S.W.: Automatic Gesture Recognition for Intelligent Human-Robot Interaction. 7th International Conference on Automatic Face and Gesture Recognition, (pp 645 - 650). (Southampton 2006).

[5] Alex Couture-Beil R.T.V., and Greg Mori: Selecting and Commanding Individual Robots in a Vision-Based Multi-Robot System. Proceedings of the 5th ACM/IEEE international conference on Human robot interaction (video session). (March 2010).

[6] Sturman, D.J.Z., D.: A Survey of Glove-based Input. IEEE Computer Graphics and Applications, (pp 30-39). (1994).

[7] S. Marcel O.B., J. Viallet, D. Collobert: Hand Gesture recognition using Input/Ouput Hidden Markov Models. FG '00 Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, (pp 456-). (Washington, DC, USA: IEEE Computer Society 2000).

[8] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. IEEE transactions on pattern analysis and machine, Vol. 19, No. 7. (July 1997).

[9] Corradini A.: Real-Time Gesture Recognition by Means of Hybrid Recognizers. GW '01 Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction, (pp 34-46). (London, UK 2001).

[10] A. Ramamoorthya N.V., S. Chaudhurya, S. Banerjeeb: Recognition of dynamic hand gestures. Pattern Recognition 36:2069 – 2081(2003).

[11] S. Waldherr R.R., and S. Thrun: A Gesture Based Interface for Human-Robot Interaction. Autonomous Robots 9:151-173(2000).

[12] T. Axenbeck M.B., S. Behnke, and W. Burgard: Recognizing complex, parameterized gestures from monocular image sequences. Proc. of the IEEE/RSJ Int. Conf. on Humanoid Robots (Humanoids). (2008).

[13] F. Faber M.B., C. Eppner, and A. Gorog: The Humanoid Museum Tour Guide Robotinho. (2009).

[14] Helman Stern J.W., and Yael Edan: A Method for Selection of Optimal Hand Gesture Vocabularies. (Springer Berlin / Heidelberg 2009).

[15] Thi-Thanh-Hai Tran, Thi-Thanh-Mai Nguyen : Invariant lighting hand posture classification, (pp 827-831). IEEE international conference on Progress in Informatics and Computing (PIC). (December 2010)

[16] P. Viola, M. Jones. Robust Real-time Object Detection: Cambridge Research Laboratory Technical Report Series; 2001.