IEEE Computer Society Order Number E5621
BMS Part Number CFP1503I-ART
ISBN 978-1-4673-8013-3

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 Lisa O'Conner
Cover Design by Mark Bartosik

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

2015 Seventh International Conference on Knowledge and Systems Engineering

KSE 2015

Table of Contents

Long Papers

First International Workshop on Pattern Recognition for Multimedia Content Analysis: PR4MCA 2015

2015 Seventh International Conference on Knowledge and Systems Engineering

Using hand postures for interacting with assistant robot in library

Van-Toi Nguyen*†‡, Thanh-Hai Tran*, Thi-Lan Le*, Rémy Mullot† and Vincent Courboulay†
*International Research Institute MICA, HUST-CNRS/UMI-2954-GRENOBLE INP
and Hanoi University of Science & Technology, Vietnam
Email: Van-Toi.Nguyen; Thanh-Hai.Tran;Thi-Lan.Le@mica.edu.vn
† L3i Laboratory, the University of La Rochelle, France
Email: remy.mullot@univ-larochelle.fr; vcourbou@univ-lr.fr
‡The University of Information and Communication Technology under Thai Nguyen University, Vietnam

***Abstract*—Visual interpretation of hand gesture for human-system interaction in general and human-robot interaction in particular is becoming a hot topic in computer vision and robotics fields. Hand gestures provide very intuitive and efficient means and enhance the flexibility in communication. Even a number of works have been proposed for hand gesture recognition, the use of these works for real human-robot interaction is still limited. Based on our previous works for hand detection and hand gesture recognition, we have built a fully automatic hand gesture recognition system and have applied it in a human-robot interaction application: service robot in library. In this paper, we describe in detail this application from the user requirement analysis to system deployment and experiments with end-users.**

I. Introduction

While robots were initially used in repetitive tasks where all human direction is given a prior, they are becoming involved in increasingly more complex and less structured tasks and activities, including interaction with human required to complete those tasks. With this reason, Human - Robot Interaction (HRI) becomes a hot topic that attracts the interests of not only academic but also popular culture researchers. They want to study on how humans interacts with the robot, and how to design and implement robot systems that are able to accomplish the interaction tasks in human environment in a *direct, safe and effective* manner.

Ideally, human - robot interaction should approach human-human interaction which composes of many forms of communication. In early day of robotics, researchers have been attempting to make them understand human speech. But in the last years, researchers try to introduce the other means of human-human interaction to the field of HRI: hand gesture. Human hand gestures are means of non-verbal interaction among people. Hand gestures have been shown to be an *intuitive and efficient* manner of communication. To make robot be able to communicate with human by hand gesture, it is essential that robot needs to be equipped with a camera to see human hand gesture and with *a brain* to analyze and understand which hand gesture in the captured image.

Hand gestures can be divided into two types: static hand posture and dynamic hand gesture. A static hand posture is described by only one pose of the hand while a dynamic hand gesture is defined by a sequence of hand postures varying in time. Hand waving is an example of dynamic hand gesture. While a static hand posture can be represented in spatial domain, a dynamic hand gesture has to be represented in temporal-spatial domain. Our work focuses on hand postures recognition. Based on our previous work for hand detection [1] and hand posture recognition [3], we have built a fully automatic hand posture recognition. In this paper we introduce a new application of human-robot interaction based on hand gesture: library service robot. The remaining of this paper is organized as follows. Section II presents briefly some hand gesture-based human-robot interaction. In Section III, we introduce the library service robot application. Section IV analyze our experiments.

II. Related works

Concerning the use of hand gesture in human-robot interaction, several research groups have applied hand gestures as a communication manner [4], [5], [6].

The Institute of Computer Design and Fault Tolerance developed a robot assistant, called Albert, that can be controlled by hand gesture and speech [4]. The robot is equipped with a color stereo camera allowing to recognize six types of hand gestures in order to stop the robot action in progress, to answer yes/no, to point to an object of interest or to explain grasp situation. The authors have validated their recognition algorithm but not experimented the communication between human and robot in a real situation.

The Faculty of Technology, Applied Computer Science, Bielefeld University developed the robot Biron as a companion of human in his household environment [5]. This robot can communicate with user through mono-manual pointing gestures. As designed to be a companion of human, for the first time, the user must give robot the tour through his private home so to familiarize it with its new habitat. The human points to the locations and objects of interest to robot. In the communication with human, only one type of hand gesture has been used.

The Institute for Computer Science of University of Bonn has a long history in developing generations of humanoid robot that can play soccer in the RoboCup Humanoid League and perform domestic tasks in RoboCup@Home League. In [6] two types of hand gesture showing and pointing in order to draw the robot's attention to a particular object have been studied. A Time-of-Light camera has been used for perceiving

978-1-4673-8013-3/15 $31.00 © 2015 IEEE
DOI 10.1109/KSE.2015.18

gestures. For pointing gestures, the pointing direction is estimated and matched with objects in the robot's environment. In the case of showing gesture, the robot tries to extract and visually recognize the shown object. One of disadvantage of this approach is the use of range camera that is not popular and quite expensive.

III. Service robot in library

A. Overview

This section presents an application of human-robot interaction using hand postures. The idea is to demonstrate that hand posture recognition could be applied successfully in real situations. In our work, the communication between human and robot takes place in a library environment where robot plays the role of a librarian. We design a set of hand postures and map them to a set of commands to control the robot in library. In the following, we will describe in more detail how we build this application and validate it.

B. Investigation of library environment and end user requirements

We have visited the Ta Quang Buu Library of Hanoi University of Science and Technology to understand the organization of a library as well as borrowing and returning book activities. The library spans on several floors of the building and is organized into specific rooms such as storing room, reading room or borrowing room. The users are mainly students. Books and users have been managed through 1D barcode management system. Every year, new students will be provided with an account for using library services. This account will be created based on the student profile including his/her student card number and the barcode onside the card. The library provides different station machines allowing students to lookup:

- User information (including user account information and list of books borrowed by this user) by entering a barcode number on the student card.
- Information of a borrowed book by entering the barcode ID of this book or search in the borrowed list.

The library used the library database management software named VTLS (http://www.vtls.com/) to manage the users, books, as well as book borrowing and returning information. When a user wants to get into the library, he/she needs to show his/her student card to a receptionist of the library. In reading rooms, the user can lookup information using OPAC interface (a module of VTLS) on a station machine. In borrow rooms, the user brings books that they want to borrow to the librarian. According to librarians as well as users, the most frequent and important query is the one about overdue books.

C. Definition of human robot interaction scenarios

After investigating the library and its activities, we propose to simulate the library Ta Quang Buu by implementing a room (that plays the role of reading and borrowing room) inside the showroom of MICA Institute. The simulated library is a room of size 3m x 3m in which we equip some tables, chairs, bookshelves. All are similar to a reading room in the library so that the human can feel as in a real library. In this context, the scenarios are played by two actors: a human and an assistant robot in the library. We focus on some general and basic requirements that a human needs and often uses in the library such as search for user information or book.

To define interaction scenarios, we invent situations and assign roles to human and robot as follows (see Fig.1):

1) *User* comes and stands in front of the robot.
2) *Robot* captures photos of user face and recognizes him.
    a) If the robot can not recognize the user, go to (3).
    b) If the user is recognized successfully, go to (4).
3) *Robot* considers as the user has not been registered before. The robot will say by synthetic voice: *"Hello, you are not registered. Please do the registration first"*.
4) *Robot* says : *"Hello, welcome to the library! I'm ready for your consultation"*. The robot searches overdue books borrowed by this user and reminds him by displaying all overdue books on the screen.
5) Once getting login, user can trigger the robot actions by hand gestures to:
    a) Ask for user information
    b) Lookup borrowed books, select a book
    c) Listen summary of a selected book
    d) Stop interaction with the robot

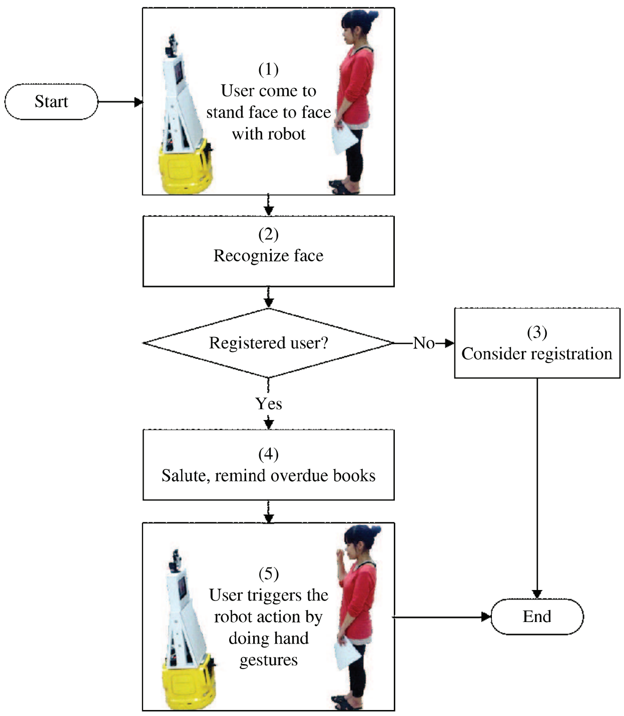

Fig. 1: The activity diagram of the system

D. Design of postural command vocabulary

With the above scenario, we design a set of hand postures and map them to set of commands for human-robot interaction. It covers two groups of commands: Interface Controlling and Consultation. Figure 2 shows the commands with corresponding hand postures.

- Interface controlling: it consists of four commands: back, next, next page and end (exit).
- Consultation: it consists of three commands: open user info, open book info, read book summary.

E. Hand detection and posture recognition

A real application of human robot interaction requires a good recognition accuracy and low computational time to be able to work in complex environments with cluttered background and view point changes. In our previous works, we have proposed a hand detection method based on Internal Haar-like feature [1] and a hand posture recognition method based on kernel descriptor [2], [3]. Our detection method allows to avoid bad effects of background changing while the hand posture recognition method is robust to rotation and scale change.

Based on these works, we have built an automatic hand posture recognition system that takes an image as input and returns the label of the hand posutre performing in the image. The system composes of two main modules: Hand detection and Hand posture recognition. The hand posture recognition consists of two steps: hand representation and hand posture classification. The overall hand detection and recognition framework is shown in Fig.3. We describe briefly these modules. The detailed information can be found in our previous papers [1], [3].

- *Hand detection*: This module scans the whole image using window sliding technique. Each scanned window will be classified as hand or non-hand region using Internal Haarlike feature and Cascaded Adaboost Classifier.
- *Hand representation*: This step takes a hand region image (hereafter called shortly as image) as input and results a descriptor of the candidate hand. It is composed of three sub-steps:
  - Pixel-level feature extraction: At this level, we compute gradient vector for each pixel of the image.
  - Patch-level feature extraction: At this level, we firstly have to generate a set of patches then compute patch-level features. Different from [7], depending on image resolution, we create patches with adaptive size instead of fixed size. This ensures the number of patches to be considered unchanged. In addition, it makes the patch descriptor more robust to scale change. For each patch, we compute patch features as presented following. Given an image patch, we compute a gradient descriptor based on the original idea proposed in [7]. However, unlike [7], we compute first the dominant orientation of the patch then normalize all gradient vectors to this orientation. This normalization is done inside the gradient kernel allowing the descriptor invariant to rotation.
  - Image-level feature extraction: At this step, we propose a modification w.r.t to [7]: To combine patch features, we propose a pyramid structure specific to hand postures instead of general pyramid structure. This makes the descriptor more suitable for hand representation. Given an image, we will build the final representation based on features extracted from lower levels using efficient match kernels (EMK) proposed in [7]. First, we have to compute feature vector for each cell of hand pyramid structure, then concatenate them into a final descriptor.
- *Hand posture classification*: Once the hand is represented by a descriptor vector, we apply Multi-class SVM in order to identify hand posture.

F. Deployment of the proposed method on robot for human-robot interaction

To deploy such method for hand posture recognition, we integrate the modules of hand detection and hand posture recognition on a robot. However, these modules are designed to take still images as input. For a real application, the camera captures consecutively frames then the question is how to make decision of recognized command at certain time in order to control the robot.

To deal with this question, several methods require the user to start controlling the robot by raising his hand, keep the posture in a certain time (normally from one to two seconds) and finish his command by putting his hand down. By this way, the system can determine more easily the period of playing hand postures then follow a voting scheme on frame by frame recognition results. This approach does not allow the user to make consecutive commands without putting down the hand. In addition, because the command is only executed after segmentation and recognition, the response could be slow.

We propose a strategy that detects and recognizes hand posture every frame and makes the decision based on a number of consecutive similar recognized hand postures and the duration between them (Fig.4). This method avoids the above mentioned drawbacks and gives satisfied results as shown in experiments section.

Suppose at time k, the system captures frame F_k and does hand detection and posture recognition. If the posture P is recognized, we confirm a command corresponding to the posture P to control the Library Management System and if the following conditions are satisfied:

1) The system detects the same hand posture P at fpc frames before $F_{k_1}, F_{k_2}, ...F_{k_{fpc}}$
2) The system detects nothing at remaining frames counting from F_{k_1} to F_k
3) The number of frames between every couple of frames $F_{k_i}, F_{k_{i+1}}$ dees not exceed fpr frames.

After sending a command to control the Library Management System, we reset parameters to start the process of the

Interface Controlling				Consultation		
Back	Next	End	Next Page	Open User Info	Open Book Info	Read Book Summary

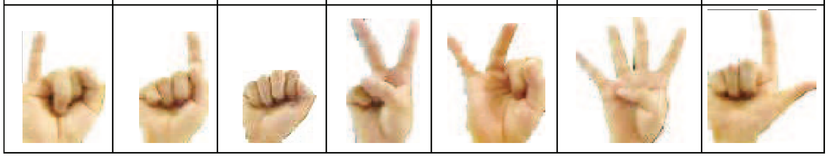

Fig. 2: Set of postural commands

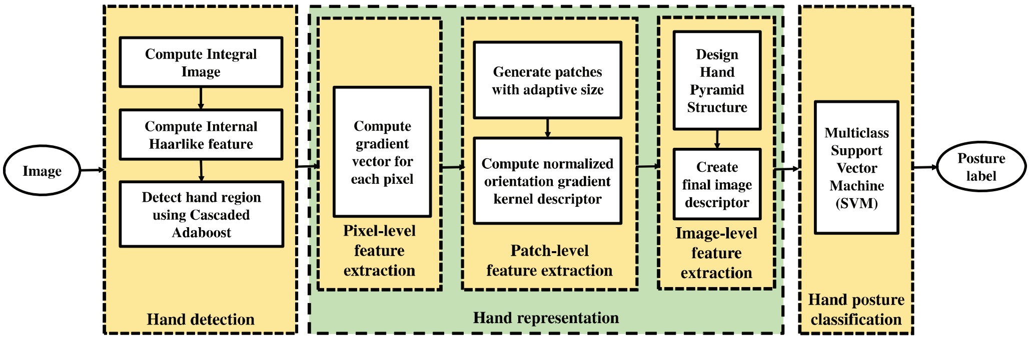

Fig. 3: The proposed framework of hand posture recognition.

upcoming command. Fig.5 visualizes how to make a decision in our deployment.

IV. Experiments

The aim of experiments is to evaluate the performance of the hand posture-based human-robot interaction in library context. The robot used for testing is a PCbot 914. It is like PC under the form of robot with 1 Gbyte of RAM DDR 2. We have mounted a frame to keep a small monitor and a camera on the robot at a height convenient for communicating with human. For evaluation, the robot is put at a corner of the simulated room, neon-lighting condition and office background. Beside, we had developed a simulated library management system with main functionalities such as VTLS. In addition, as presented section III-C, face recognition is used to login to the library management system. The face detection and recognition module has been developed in our previous works [9].

We invite 10 subjects to play the pre-defined scenario. The data coming from the first five subjects are used for training the system while the remaining subjects participate to evaluate the system. Each subject is required to trigger any command using hand postures in order to control the robot so as one posture could appear several times. We collect all the frames the system processed. The ground truth is created manually on the original captured data without any information of the results generated by the system. We compare the results with the ground truth to evaluate the performance of the system.

Table I shows the confusion matrix of the result. We add the column "Other" for the missed recognitions and the row "Other" for the false recognition. The system obtains very good performance on most of hand postures except the "Next page". In the overall result, the average Precision is 83%, the average Recall is 91%, and the average F-score is 86%. We observe that the hand pyramid help distinguish hand postures very similar with different open/close fingers. We can see in Figure 3, the postures for "Next Page" and "Open User info" are very similar except for the fist and the particular fingers involved in the postural command. As you can see in the confusion matrix in Table I, our system does not make confusions between these two hand postures, though. The confusion matrix shows a perfect performance for the "Open User info". We see a relatively degraded performance for the "Next Page" command. Fig. 6 illustrates an example of wrong recognition. The hand is well detected. The hand posture recognition module, however, makes a wrong decision because of the participation of background information in the hand bounding box that makes the gradient of this rectangle region is similar to the gradient of "Open bookinfo" posture regions. Removing the unexpected effects of background on

TABLE I: The confusion matrix of action recognition

Action	Back	Next	Open book info	End	Open user info	Next page	Read book sum-mary	Other
Back	30	0	0	0	0	0	0	0
Next	0	30	0	0	0	0	0	0
Open book info	0	0	30	0	0	0	0	0
End	0	0	0	30	0	0	0	0
Open user info	0	0	0	0	30	0	0	0
Next page	0	0	10	0	0	10	0	10
Read book summary	0	0	0	0	0	0	30	0
Other	0	0	0	0	0	20	0	0

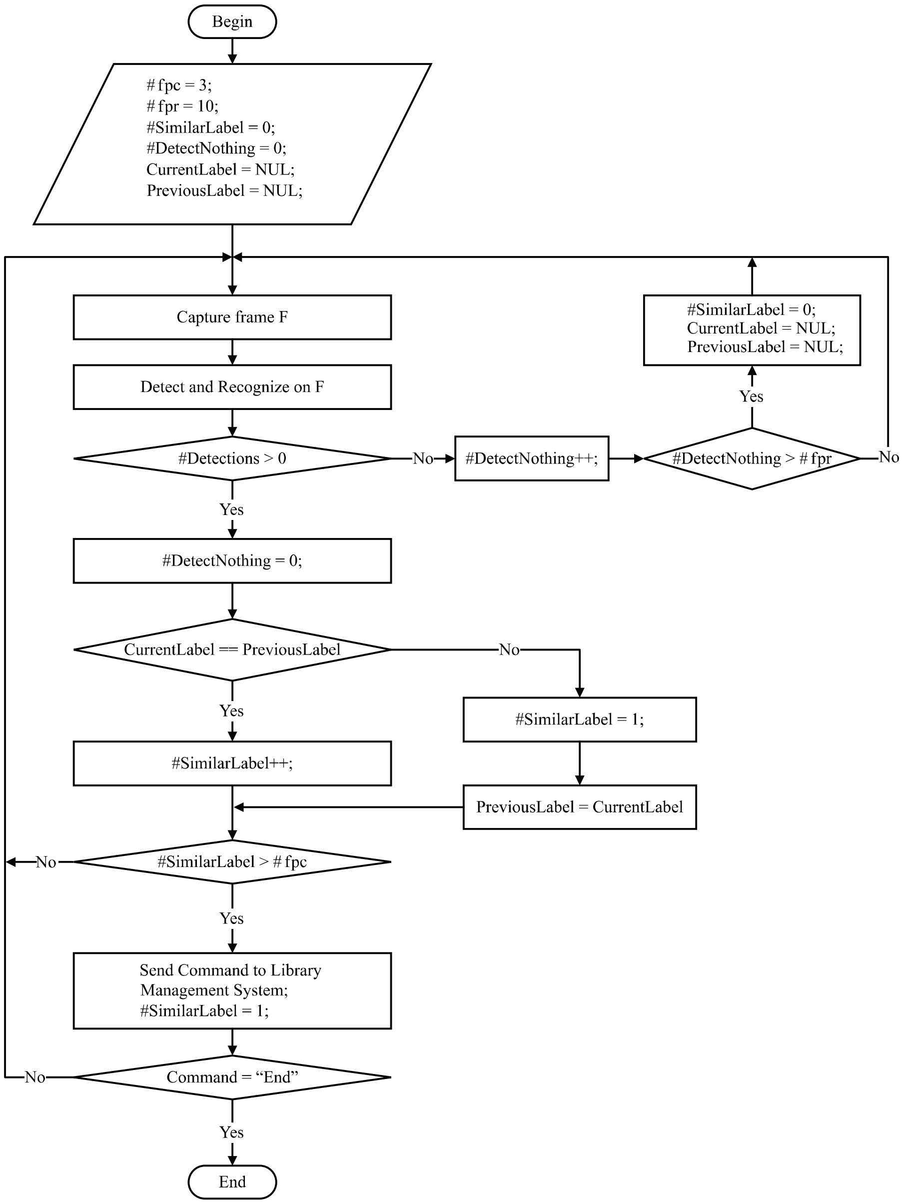

Fig. 4: System deployment.

hand posture recognition is a further topic need to research.

Fig.7 shows an example image captured from working session. The user makes the command "Open Book Info", the system recognized and goes to the book info page.

V. Conclusion and Feature works

This paper shows a fully automatic hand posture recognition system, integrated successfully on a robot for human-robot interaction application in library context. Extensive experiments show that the proposed system works well in real situation and responses in a very satisfied time duration. We observed also that the people feel very comfortable and funny to use hand gesture to communicate with the system. In the future works, we will combine hand gestures with other modality such as speech to make the interaction more natural and efficient. The paper shows a specific context but the framework could be extended for more generic application.

Acknowledgment

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number FWO.102.2013.08

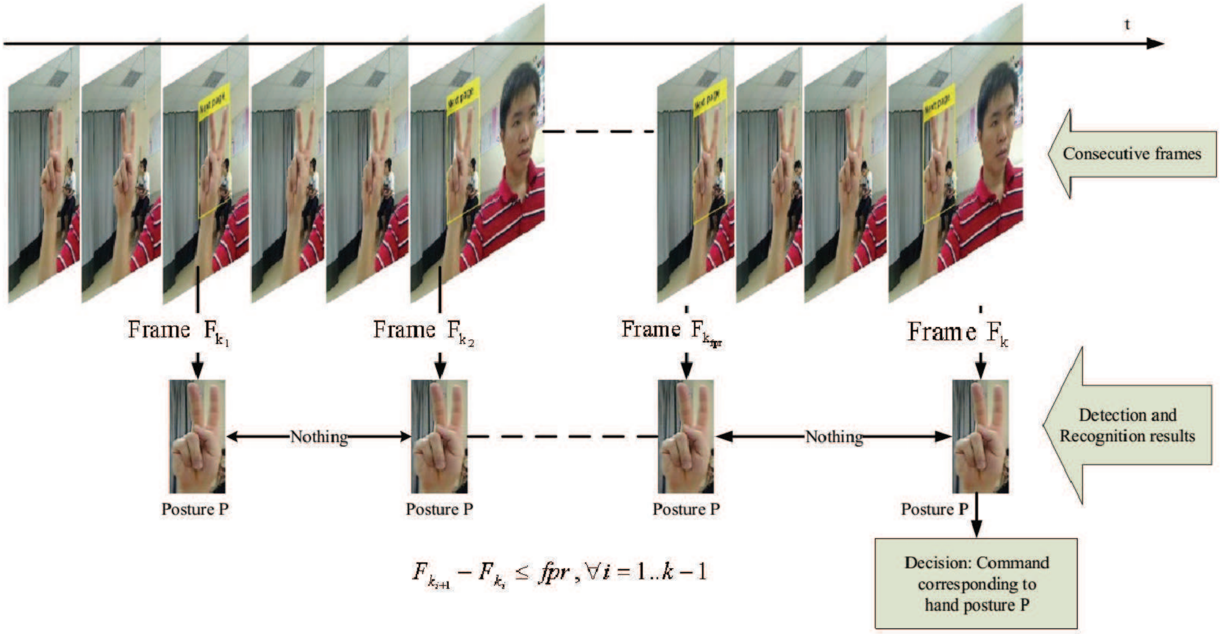

Fig. 5: System deployment: making decision.

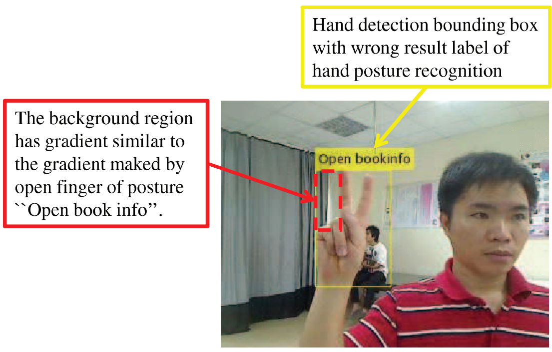

Fig. 6: An example of wrong recognition. A "Next page" posture is recognized as "Open bookinfo".

Fig. 7: An example image captured from working session.

References

[1] V.T. Nguyen, T.L. Le, T.T.H. Tran, R. Mullot, and V. Courboulay, "A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier," In Proceeding of The International Conference on Communications and Electronics (ICCE), 2012.

[2] V.T. Nguyen, T.L. Le, T.T.H. Tran, R. Mullot, and V. Courboulay, "Hand Posture Recognition Using Kernel Descriptor," In Proceeding of the International Conference on Intelligent Human-Computer Interaction (iHCI), 2014.

[3] V.T. Nguyen, T.L. Le, T.T.H. Tran, R. Mullot, and V. Courboulay, "A New Hand Representation Based on Kernels for Hand Posture Recognition," In Proceeding of The Eleventh IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015), Ljubljana, Slovenia, 2015.

[4] O. Rogalla, M. Ehrenmann, R. Zollner, R. Becher, and R. Dillmann, "Using gesture and speech control for commanding a robot assistant, Proceedings," 11th IEEE International Workshop on Robot and Human Interactive Communication, 2002, pp. 454-459.

[5] J.F. Maas, T. Spexard,J. Fritsch, B. Wrede, and G. Sagerer, "Biron, what's the topic? a multi-modal topic tracker for improved human-robot interaction," The 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN 2006), 2006, pp. 26-32.

[6] D. Droeschel, J. Stuckler, D. Holz, and S. Behnke, "Towards joint attention for a domestic service robot-person awareness and gesture recognition using time-of-flight cameras," 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 1205–1210, 2011.

[7] L. Bo, X. Ren, and D. Fox, "Kernel Descriptors for Visual Recognition," NIPS, 2010, pp. 1-9.

[8] N.H Dardas and N.D Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," IEEE Transactions on Instrumentation and Measurement (TIM), pp. 3592-3607, 2011.

[9] T.T.H. Tran, "Face recognition from video under uncontrolled lighting condition," In Proceeding of the 12th IEEE International Symposium on Signal Processing and Information Technology (ISPPIT), 2012

IEEE KSE 2015

The 1st International Workshop on Pattern Recognition for Multimedia Content Analysis

PR4MCA 2015

In conjunction with The 7th International Conference on Knowledge and Systems Engineering

October 8 - 10, 2015 - Ho Chi Minh, Vietnam

PR4MCA - BEST PAPER AWARD

presented to

Van-Toi Nguyen, Thanh-Hai Tran, Thi-Lan Le,
Mullot Rémy and Vincent Courboulay

for the paper

Using hand postures for interacting with assistant robot in library

On behalf of PR4MCA Organizers

Dung Duc Nguyen