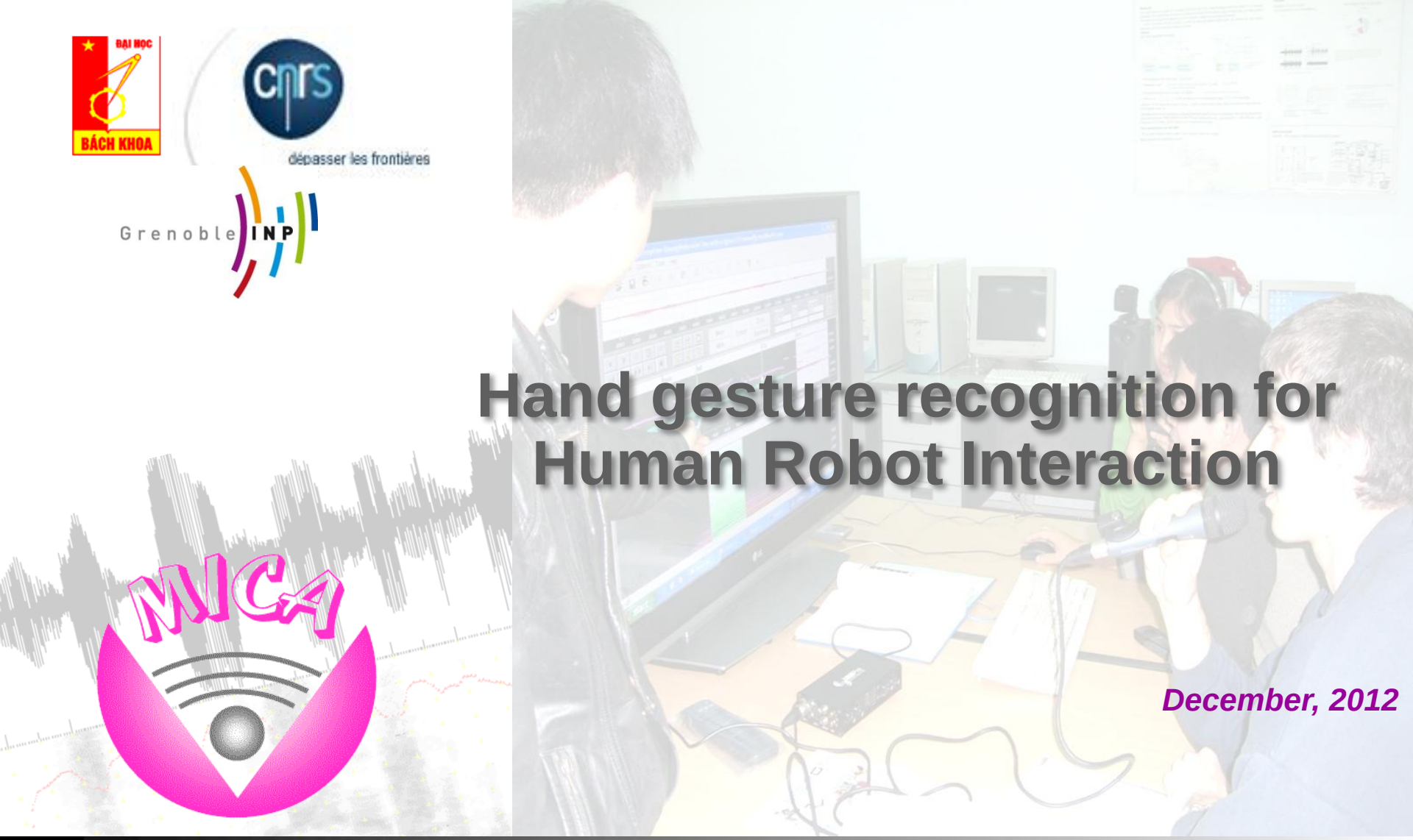


Hand gesture recognition for Human Robot Interaction

December, 2012

International Research Institute MICA
Multimedia, Information, Communication & Applications
UMI 2954

Hanoi University of Science and Technology
1 Dai Co Viet - Hanoi - Vietnam

Outline

- Introduction
- Objectives and proposed solution
- Experiments and results
- Conclusions and future works

Introduction

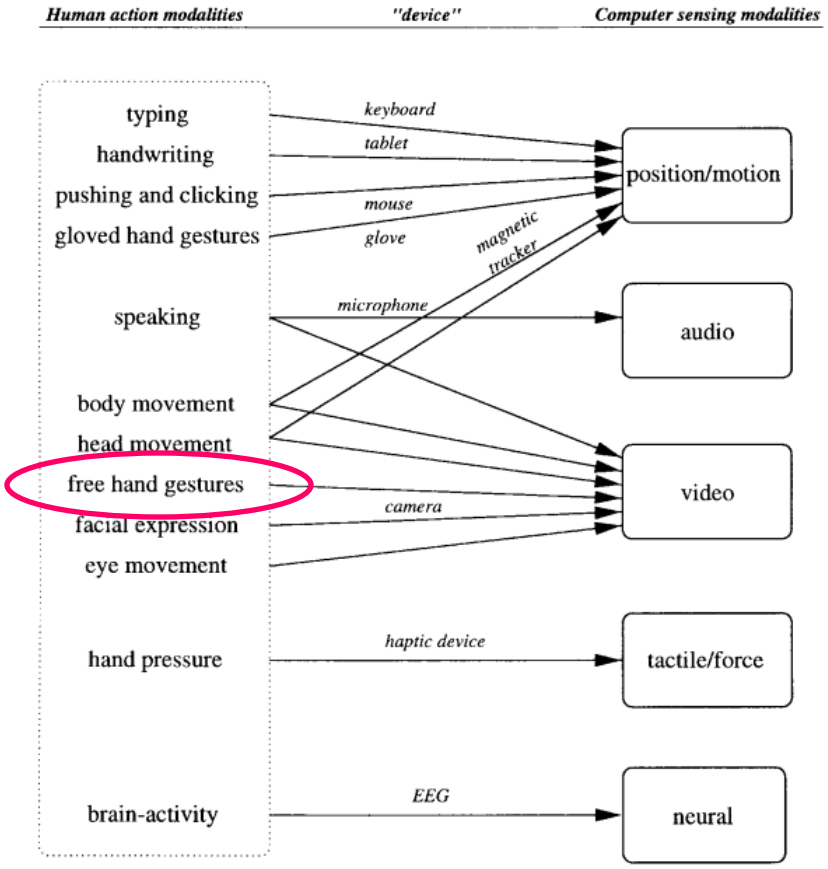
- At MICA

- We would like to build applications using Robot as Guide in different environment (i.e Library, Museum)

Introduction

How does human communicate with robot ?

Intuitive, Natural, Efficient



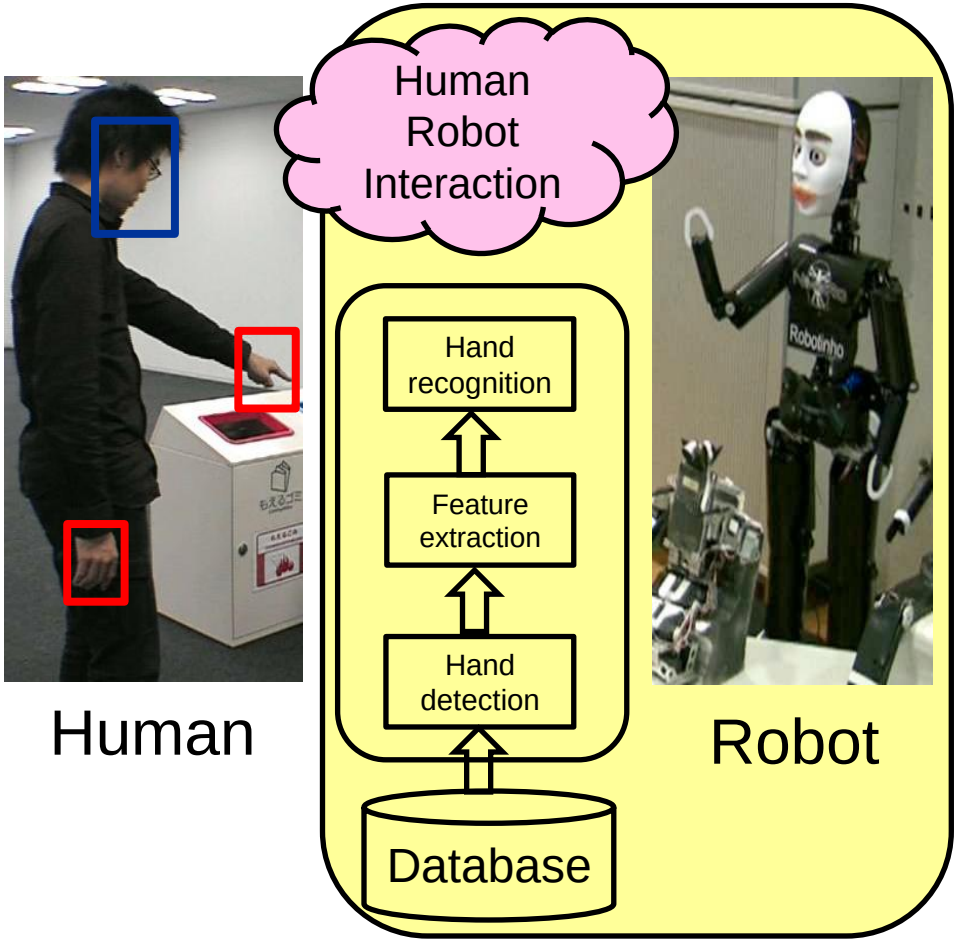
Hand gestures have been shown to be an intuitive and efficient mean

- **to express an idea**
- **to control something**

Objectives

- **A vocabulary of hand gestures** needs to be defined, and a gesture based protocol of communication should be understood by both human and robot.
- **A system for hand gesture recognition** must be built so that it could be integrated in robot for an automated interaction.

HRI by Hand Gesture



State of the art

- **A vocabulary of hand gestures** needs to be defined, and a gesture based protocol of communication should be understood by both human and robot.
- **A system for hand gesture recognition** must be built so that it could be integrated in robot for an automated interaction.

- **About more than 10 public databases of hand gestures**
- **But:**
  - The methodology for designing and building a hand gesture database has not been mentioned yet.
  - It's imposed for human without considering if they do this in a comfortable manner or not
  - We need redefining a gesture set for each specific application.
  - There is not exist a hand gestures database for Vietnamese
- **So: It should be useful to study and to design a common set of hand gestures that could be used for general context.**

State of the art

- **A vocabulary of hand gestures** needs to be defined, and a gesture based protocol of communication should be understood by both human and robot.
- **A system for hand gesture recognition** must be built so that it could be integrated in robot for an automated interaction.

- **Feature Extraction**: Haarlike, SIFT, Ridge, Blob, etc.
- **Classification Method**: Cascaded Adaboost, SVM, Neural Network, etc.
- **But:** There is no exact answer for the question: which method is the best for hand gestures recognition?
- **For real time applications**, Haar-like features and Cascaded Adaboost Classifer give the good performance in term of computational time and precision

Proposed solution

Needs

- **A vocabulary of hand gestures** needs to be defined, and a gesture based protocol of communication should be understood by both human and robot.
- **A system for hand gesture recognition** must be built so that it could be integrated in robot for an automated interaction.

Propositions

- **A framework for designing hand gesture set for Vietnamese** uses the Wizard of Oz technique
- **A system for hand gesture recognition** uses the Haar-like features and the Cascaded Adaboost classifier

A framework for designing hand gesture set for Vietnamese using the Wizard of Oz technique

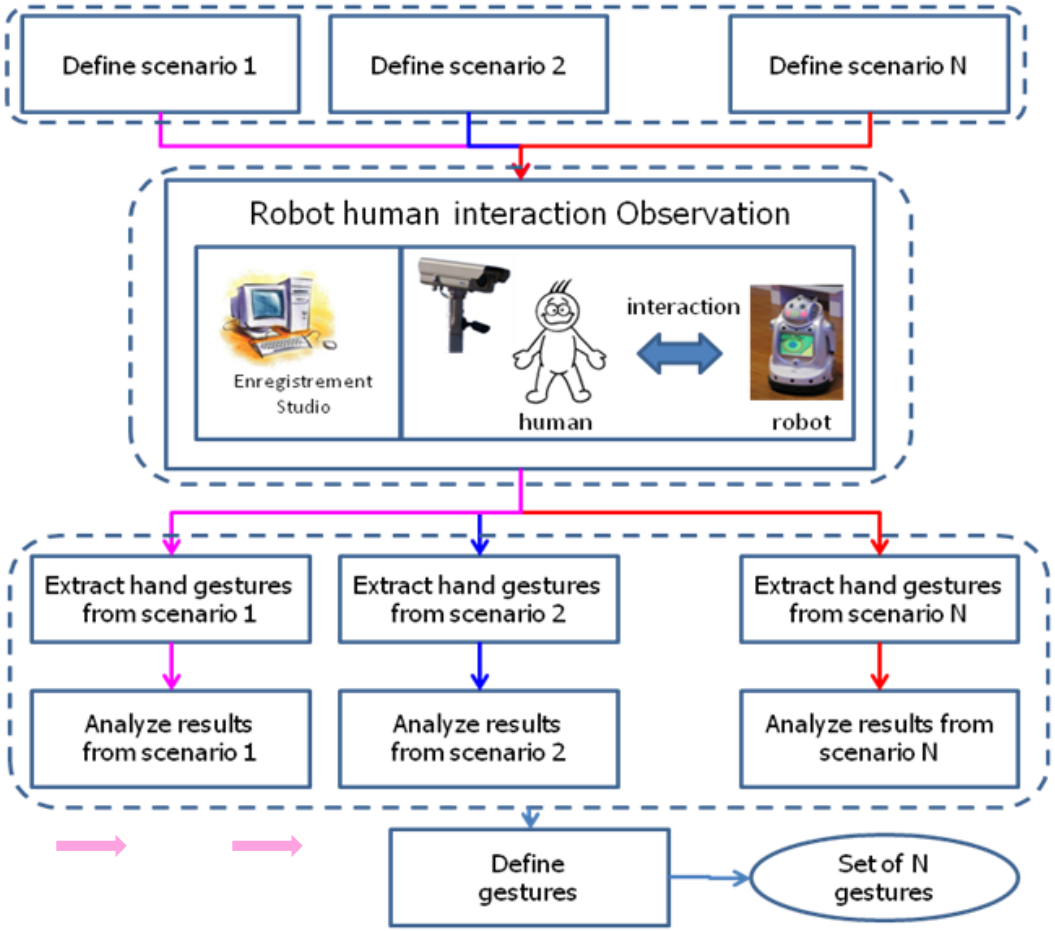
Hand gesture vocabulary design

Task 1: Define interaction scenarios

Task 2: Observe (record) robot human interaction

Task 3: Extract hand gestures and analysis

Task 4: Define gestures set



Framework of designing hand gesture vocabulary

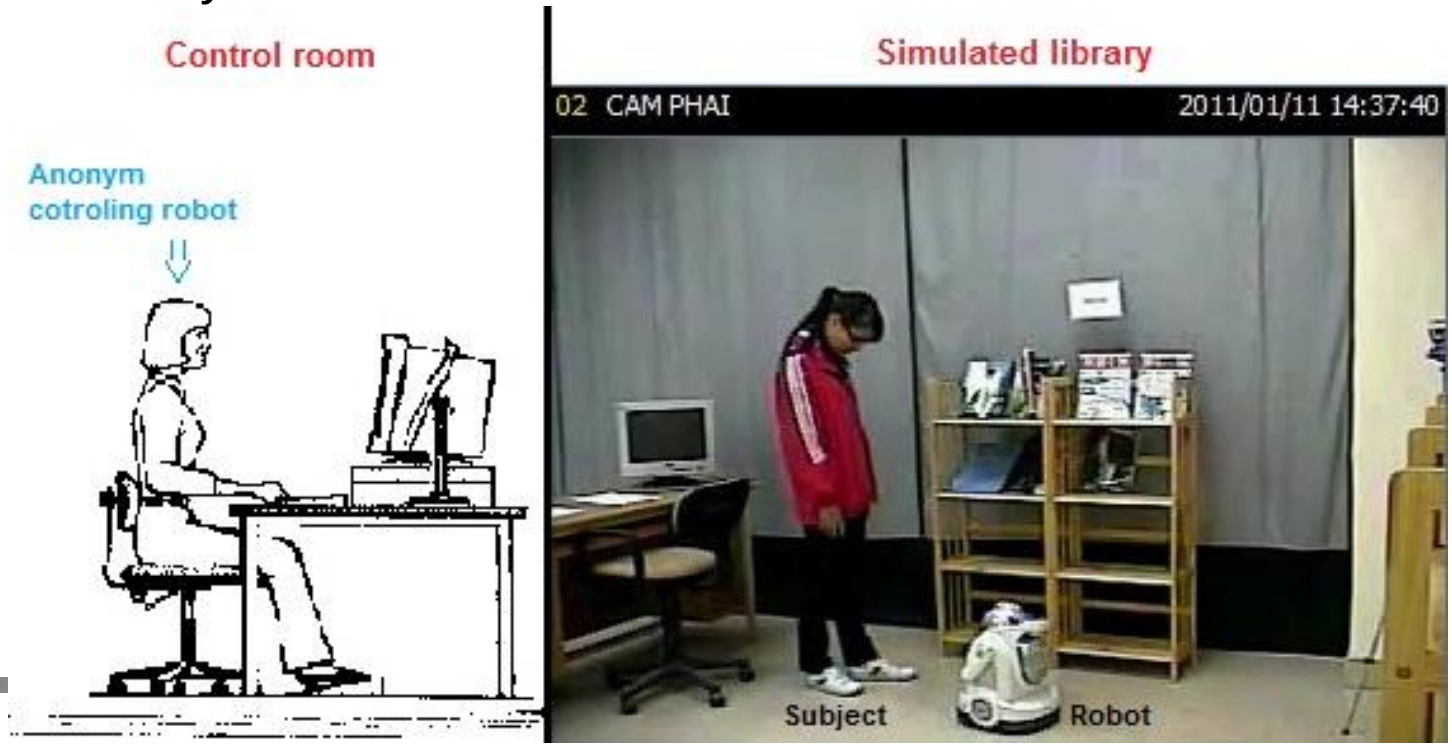
Task 1: Definition of HRI scenarios

- **Define a serie of HRI scenarios in a simulated library context:**

- **Study behaviors of human interacting with robot in the most 5 common situations:**
  1. Call robot
  2. Point to something for a service
  3. Agree with robot's answer
  4. Disagree with robot's answer
  5. Finish the human – robot interaction.

Task 2: HRI observation

- **We use the Wizard of Oz technique** to obtain the natural HRI.
  - We say to all subjects that we would like *to test the robot's abilities*
  - All subjects do not know that the robot is controlled by an anonym technician in another room.



Task 2: HRI observation

- **A multimodal corpus (video/audio) was built with:**
  - 22 native Vietnamese people (11 males + 11 females); the mean age: 23.
  - Using 3 cameras

- **All people are asked to:**
  - Play 5 different predefined scenarios using voice, hand gestures.
  - Play 2 times all the defined scenarios, yielding 66 video files (22 subjects x 3 cameras).
- **After selecting and editing, we have obtained 850 clips, each presents only one hand gesture per scenario**

Task 3: Hand gestures extraction and analysis

- **The analysis should answer to the following questions:**
  - Which gestures are used in each scenario?
  - How are gestures characterized?

- **Some analysis results: In human – robot interaction:**
  1. Vietnamese people have trend to move the hand more than when he interacts with human in order to impress the robot
  2. The time performing one gesture in HRI is longer than the one in human – human interaction
  3. For each command, several types of hand gestures are used.

Task 4: Definition of hand gestures set

The designing of a hand gesture vocabulary needs to satisfy 2 criteria:

Criteria

For Human:The comfortableness when doing it

For Robot: The recognisability when observing it

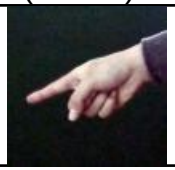
Solutions

Choosing mostly used hand gestures

Choosing distinct hand gestures

Task 4: Definition of hand gestures set

- **The hand gestures that are mostly used**

Gestures	Call (Call2)	Point (Point2)	Agree (Agree2)	Disagree (Disagree1)	Stop (Stop1)
Illustration					
Percent	92%	77%	61%	82%	96%

Problem: Hand gestures need to be distinct but Disagree and Stop are the same

Resolved by: Choose another Disagree or Stop gesture => Replace Disagree1 by Disagree2

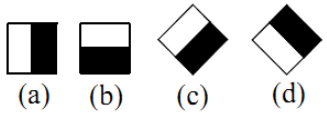
- **The hand gestures that are mostly used and distinct**

Gestures	Call (Call2)	Point (Point2)	Agree (Agree2)	Disagree (Disagree2)	Stop (Stop1)
Illustration					
Percent	92%	77%	61%	18%	96%

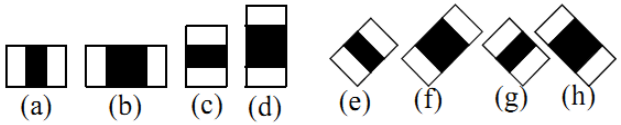
A system for hand gesture recognition using Haar-like features and Cascaded Adaboost classifier

Features and features extraction

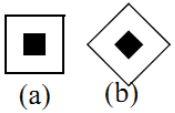
- **Proposed to use Haar-like features**
- **Haar-like features are characterized by:**
  - A corner, size, orientation
  - A value = the difference between the sum of all “white” pixel values and the one of all “black” pixel values.
- **Types of Haarlike features**
  - **Edge features**

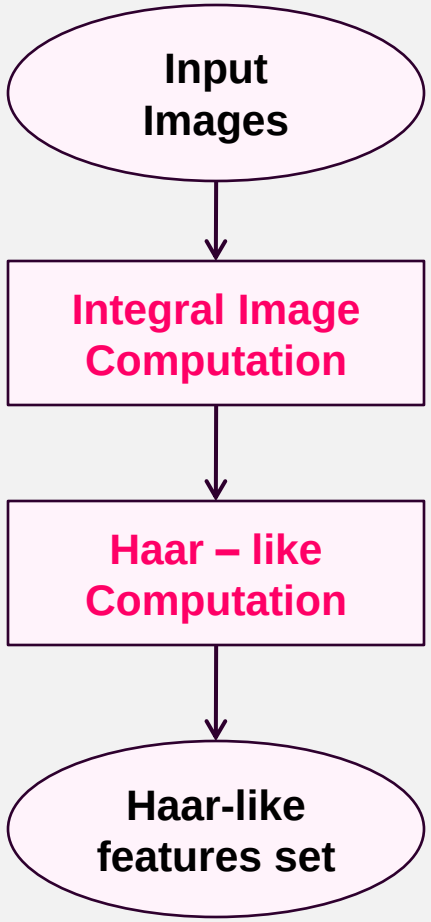


  - **Line features**



  - **Center surround features**





Computing algorithm of Haar-like features

Hand gesture classification

The number of Haar-like features are computed for one image is significantly bigger than the image size (image resolution).

For example: with an image of size 22 x 22 ~ 100.000 features

BUT

There are only some features which are significant and discriminated for posture classification

Problem:

How can we choose only the features which are significant and discriminated for posture classification?

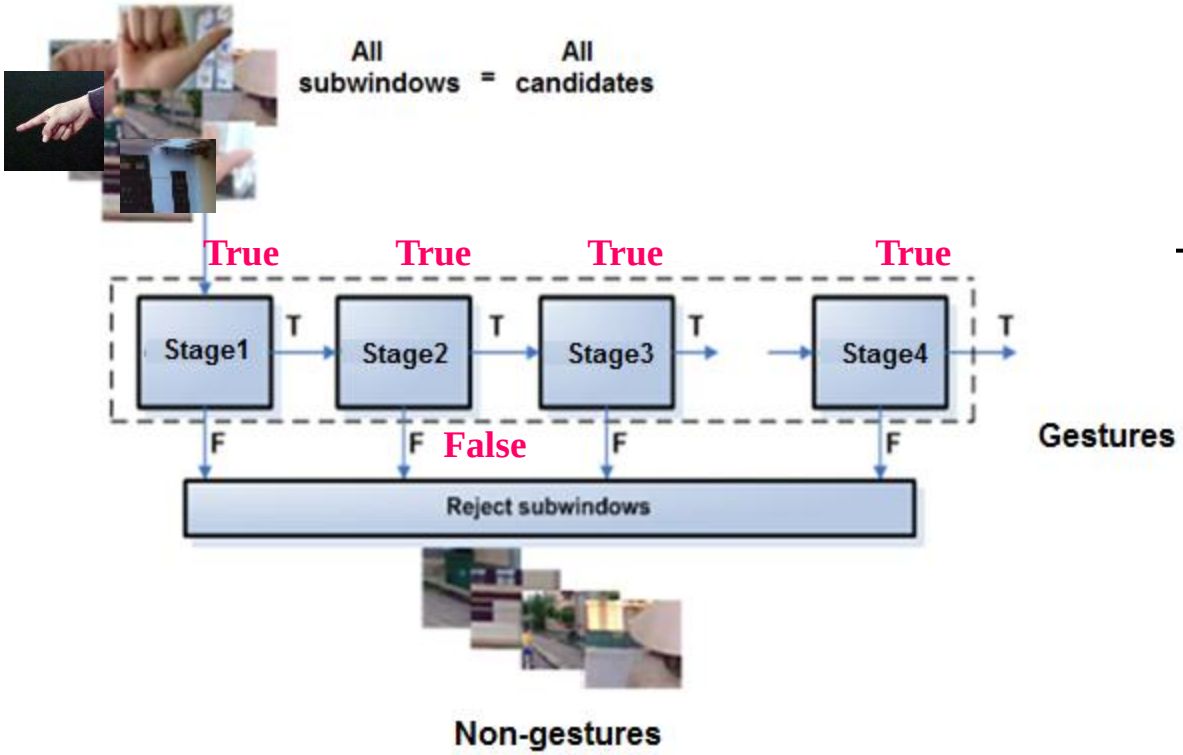
Resolved by:

The Cascaded Adaboost Classifier

Use Adaboost algorithm with only a small number of features (7 - 35 in our experiment).

Hand gesture classification

- **Cascaded Adaboost Classifier is composed of several stages**
- **Each stage is an Adaboost classifier with different max false alarms, min detection rates, and number of features**
- **A candidate will be classified into one category if it passes all stages.**



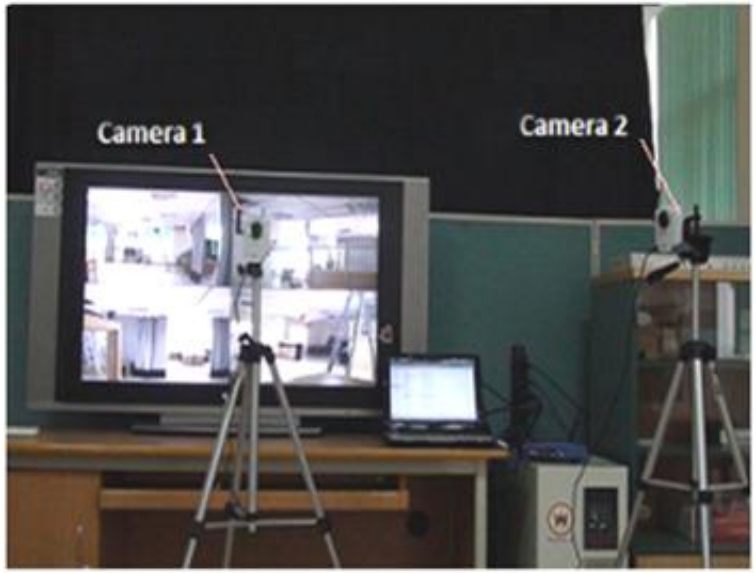
Max false alarms
Number of features

Min detection rates

Advantages:
Reject rapidly candidates which are not hand posture

Hand gesture database construction

- **Recorded database:**
  - In neon lighting condition
  - Used 2 cameras
  - With 20 Vietnamese subjects (10 females and 10 males), the age from 20 to 30 years old



- **Hand gesture database contains 1200 videos, 5 seconds per video.**
  - 3 times for each gesture
  - 20 subjects
  - 2 backgrounds (uniform and complex)
  - 5 gestures
  - 2 cameras (frontal, profile)

Experiments – Training classifiers

- **Build a system to recognize static gestures = key posture of each dynamic gestures.**
- **For each gesture, we train classifiers with:**
  - 1200 positive images (60 images/person x 20 subjects):
    - 600 images in the uniform back ground
    - 600 images in the complex back ground.
    - In the same neon light condition
  - 1500 negative images

	Gestures	Call	Point	Agree	Disagree	Stop
Positive images	Uniform background					
	Complex background					
Negative images						

Images in training database

Performance evaluation

- **Two recognition experiments on:**
  - Dependent subject: 2 subjects
  - Independent subject: 4 subjects
  - 500 positive images + 50 negative images for each subject
- **The system was evaluated by 2 criteria:**
  - **The recognition capability : recall and precision rate.**

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

	Actual class	
Predicted class	**TP** (true positive)	**FP** (false positive)
	FN (false negative)	**TN** (true negative)

  - **The computation time = the number of frames that the system can recognize per second.**

Experimental results

- **The recognition capability :**

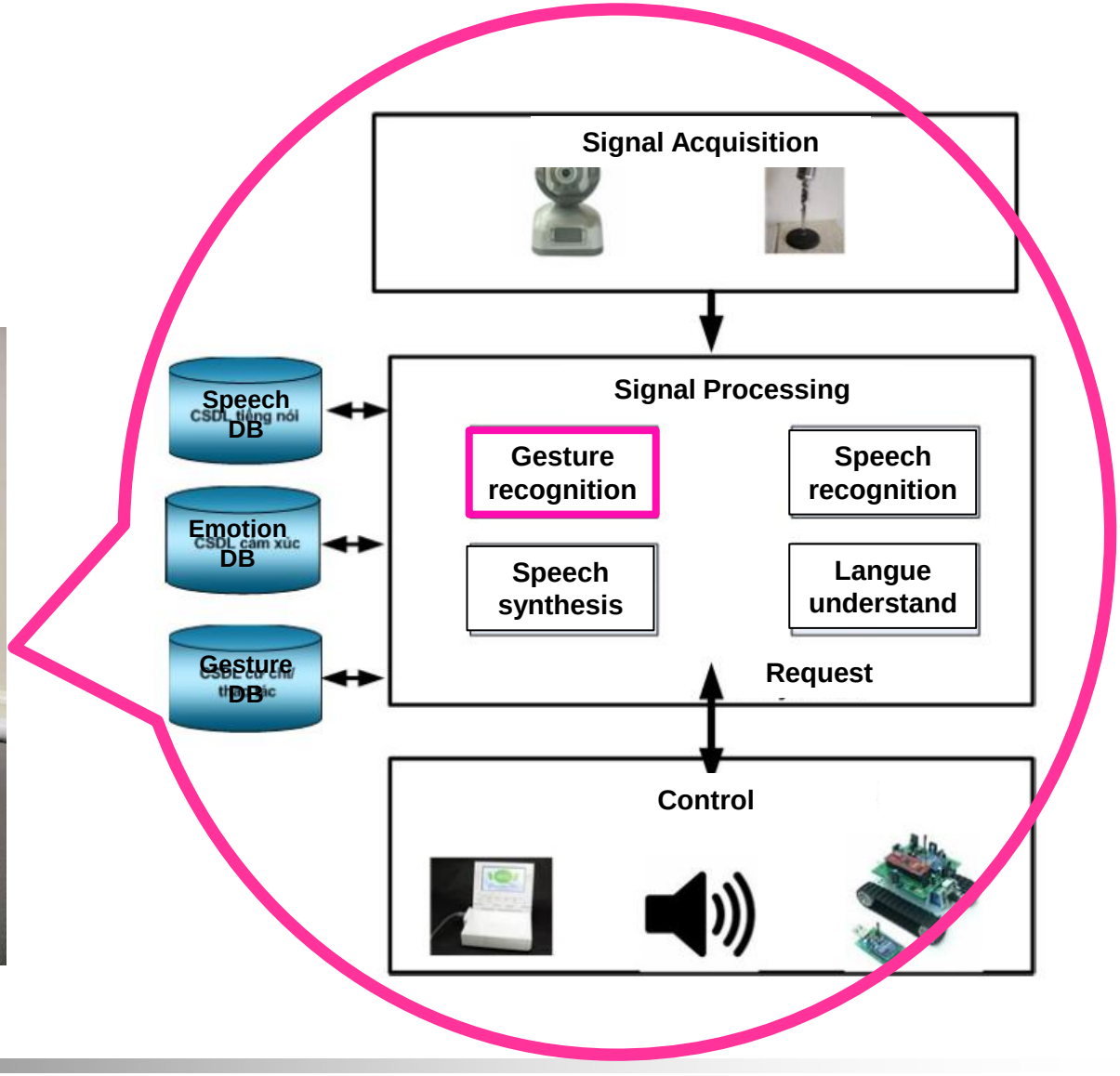
Posture	Illustration	Dependent subject experiment		Independent subject experiment	
		Recall	*Precision*	*Recall*	*Precision*
Call		89%	93%	93%	96%
Agree		67%	72%	74%	76%
Disagree		98%	96%	93%	88%
Point		92%	95%	89%	87%
Stop		95%	89%	94%	95%
Mean		88%	89%	88%	88%

- **The computation time:** is about **18 fps** on a dual core 2.66 MHz, RAM 2GB PC system.

Integration on the robot guide in the museum

- **Robot Pcbot:**

iGoLogic i3899 Mini-ITX motherboard, Intel Core 2 DUO 2 GHz, 2 Gbyte, PC3200 DDR 400MHz DIMM



Integration on the robot guide in the museum

- **Scenario: Robot is active to communicate with users**
  - Robot detects faces in its camera's field and start to communicate with the person having the biggest face.
  - The robot says *Hello* to the human by synthesis speech then proposes some services to him like presentation about the robot, presentation of objects in the museum.
  - If the robot can not detect face, human can use **Call** gesture to say hello and call the robot to come for asking a service.
  - The human can point to an object (using **Point** gesture) in the museum and ask information about this object.
  - Human expresses his attitude to the robot through hand gestures **Agree**, **Disagree**.
  - Robot gives suitable answers *Sorry* in case of disagree and Thank you in case of agree by synthesis speech.
  - When all information are provided, the user can stop the communication using **Stop** hand gesture.

Integration on the robot guide in the museum

- **Real experiments**
  - 10 participants
  - 5 hand gestures
  - Recognition rate: 77 %

Conclusions and future works

- **Our main contributions:**
  - We have studied the behavior of Vietnamese in using of hand gesture in HRI.
  - The study has been carried out through a wizard of OZ framework of 4 steps => is general and could be used for all other studies aiming finding out other interaction methods.
  - Result is a set of hand gestures (5 hand gestures):
    - commonly used in HRI applications
    - satisfying *comfortableness* and *recognisability* criteria.
- **In the future:**
  - Improve the robustess of the hand gesture recognition to the viewpoint and illumination changes