12

User, Gesture and Robot Behaviour Adaptation for Human-Robot Interaction

Md. Hasanuzzaman[1] and Haruki Ueno[2]
[1]*Department of Computer Science & Engineering, University of Dhaka,Dhaka*
[2]*National Institute of Informatics (NII), Tokyo*
[1]*Bangladesh*
[2]*Japan*

1. Introduction

Human-robot interaction has been an emerging research topic in recent year because robots are playing important roles in today's society, from factory automation to service applications to medical care and entertainment. The goal of human-robot interaction (HRI) research is to define a general human model that could lead to principles and algorithms allowing more natural and effective interaction between humans and robots. Ueno [Ueno, 2002] proposed a concept of Symbiotic Information Systems (SIS) as well as a symbiotic robotics system as one application of SIS, where humans and robots can communicate with each other in human friendly ways using speech and gesture. A Symbiotic Information System is an information system that includes human beings as an element, blends into human daily life, and is designed on the concept of symbiosis [Ueno, 2001]. Research on SIS covers a broad area, including intelligent human-machine interaction with gesture, gaze, speech, text command, etc. The objective of SIS is to allow non-expert users, who might not even be able to operate a computer keyboard, to control robots. It is therefore necessary that these robots be equipped with natural interfaces using speech and gesture.

There are several researches on human robot interaction in recent years especially focussing assistance to human. Severinson-Eklundh et. al. have developed a fetch-and-carry-robot (Cero) for motion-impaired users in the office environment [Severinson-Eklundh, 2003]. King et. al. [King, 1990] developed a 'Helpmate robot', which has already been deployed at numerous hospitals as a caregiver. Endres et. al. [Endres, 1998] developed a cleaning robot that has successfully been served in a supermarket during opening hours. Siegwart et. al. described the 'Robox' robot that worked as a tour guide during the Swiss national Exposition in 2002 [Siegwart, 2003]. Pineau et. al. described a mobile robot 'Pearl' that assists elderly people in daily living [Pineau, 2003]. Fong and Nourbakhsh [Fong, 2003] have summarized some applications of socially interactive robots. The use of intelligent robots encourages the view of the machine as a partner in communication rather than as a tool. In the near future, robots will interact closely with a group of humans in their everyday environment in the field of entertainment, recreation, health-care, nursing, etc.

Although there is no doubt that the fusion of gesture and speech allows more natural human-robot interaction, for single modality gesture recognition can be considered more reliable than speech recognition. Human voice varies from person to person, and the system

needs to take care of large number of data to recognize speech. Human speech contains three types of information: who the speaker is, what the speaker said, and how the speaker said it [Fong, 2003]. Depending on what information the robot requires, it may need to perform speaker tracking, dialogue management or even emotion analysis. Most systems are also sensitive to mis-recognition due to the environmental noise. On the other hand, gestures are expressive, meaningful body motions such as physical movements of head, face, fingers, hands or body with the intention to convey information or interact with the environment. Hand and face poses are more rigid, though its also varies little from person to person. However, humans will feel more comfortable in pointing at an object than in verbally describing its exact location. Gestures are an easy way to give geometrical information to the robot. However, gestures are varying among individuals or varying from instance to instance for a given individual. The hand shape and human skin-color are different for different persons. The gesture meanings are also different in different cultures. In human-human communications, human can adapt or learn new gestures or new users using own intelligence and contextual information. Human can also change each other behaviours based on conversation or situation. Achieving natural gesture-based interaction between human and robots, the system should be adaptable to new users, gestures and robot behaviors. This chapter includes the issues regarding new users, poses, gestures and behaviours recognition and adaptation for implementing human-robot interaction in real-time.

Adaptivity is the biological property in all creatures to survive in the biological world. It is the capability of self-modification that some agents have, which allows them to maintain a level of performance in front of environmental changes, or to improve it when confronted repeatedly with the same situation [Torras, 1995]. Gesture-based human-robot natural interaction system could be designed so that it can understand different users, their gestures, meaning of the gestures and the robot behaviours. Torras proposed robot adaptivity technique using neural learning algorithm. This method is computationally inexpensive and there is no way to encode prior knowledge about the environment to gain the efficiency. It is essential for the system to cope with the different users. A new user should be included using on-line registration process. When a user is included the user may wants to perform new gesture that is ever been used by other persons or himself/herself. In that case, the system should include the new hand poses or gestures with minimum user interaction.

In the proposed method, a frame-based knowledge model is defined for gesture interpretation and human-robot interaction. In this knowledge model, necessary frames are defined for the known users, robots, poses, gestures and robot behaviours. The system first detects a human face using a combination of template-based and feature-invariant pattern matching approaches and identifies the user using the eigenface method [Hasanuzzaman 2007]. Then, using the skin-color information of the identified user three larger skin-like regions are segmented from the YIQ color spaces, after that face and hand poses are classified by the PCA method. The system is capable of recognizing static gestures comprised of face and hand poses. It is implemented using the frame-based Software Platform for Agent and Knowledge Management (SPAK) [Ampornaramveth, 2001]. Known gestures are defined as frames in SPAK knowledge base using the combination of face and hand pose frames. If the required combination of the pose components is found then corresponding gesture frame will be activated. The system learns new users, new poses using multi-clustering approach and combines computer vision and knowledge-based approaches in order to adapt to different users, different gestures and robot behaviours.

New robot behaviour can be learned according to the generalization of multiple occurrence of same gesture with minimum user interaction. The algorithm is tested by implementing a human-robot interaction scenario with a humanoid robot "Robovie" [Kanda, 2002].

2. Related research

In this chapter we have described a vision and knowledge-based user and gesture recognition as well as adaptation system for human–robot interaction. Following subsections summarize the related works.

2.1 Overview of human-machine interaction systems

Both machines and human measure their environment through sense or input interfaces and modify their environment through expression or output interfaces. Most popular mode of human-computer or human-intelligence machine interaction is simply based on keyboards and mice. These devices are familiar but lack of naturalness and do not support remote control or telerobotics interface. Thus in recent years researchers are giving tedious pressure to find attractive and natural user interface devices. The term natural user interface is not an exact expression, but usually means an interface that is simple, easy to use and seamless as possible. Multimodal user interfaces are a strong candidate for building natural user interfaces. In multimodal approaches user can include simple keyboard and mouse with advance perception techniques like speech recognition and computer vision (gestures, gaze, etc.) as user machine interface tools.

Weimer et. al. [Weimer, 1989] described a multimodal environment that uses gesture and speech input to control a CAD system. They used 'Dataglove' to track the hand gestures and presented the objects in three-dimension onto the polarizing glasses. Yang et. al. [Yang, 1998] have implemented a camera-based face and facial features (eyes, lips and nostrils) tracker system. The system can also estimate user gaze direction and head poses. They have implemented two multimodal applications: a lip-reading system and a panoramic image viewer. The lip-reading systems improve speech recognition accuracy by using visual input to disambiguate among acoustically confusing speech elements. The panoramic image viewer uses gaze to control panning and tilting, and speech to control zooming. Perzanowski et. al. [Perzanowski, 2001] proposed multimodal human-robot interface for mobile robot. They have incorporated both natural language understanding and gesture recognition as a communication mode. They have implemented their method on a team of 'Nomad 200' and 'RWI ATRV-Jr' robots. These robots understand speech, hand gestures and input from a handheld Palm pilot to other Personal Digital Assistant (PDA).

To use the gestures in the HCI or HRI it is necessary to interpret the gestures by computer or robot. The interpretation of human gestures requires that static or dynamic modelling of the human hand, arm, face and other parts of the body that is measurable by the computers or intelligent machines. First attempt is to measure the gesture features (hand pose and/or arm joint angles and spatial positions) are by the so called glove-based devices [Sturman, 1994]. The problems regarding gloves and other interface devices can be solved using vision-based non-contract and nonverbal communication techniques. Numbers of approaches have been applied for the visual interpretation of gestures to implement human-machine interaction [Pavlovic, 1997]. Torrance [Torrance, 1994] proposed a natural language-based interface for teaching mobile robots about the names of places in an indoor environment. But due to the

lack of speech recognition his interface system still uses keyboard-based text input. Kortenkamp et. al. [Kortenkamp, 1996] have developed gesture-based human-mobile robot interface. They have used static arm poses as gestures. Stefan Waldherr et. al. proposed gesture-based interface for human and service robot interaction [Waldherr, 2000]. They combined template-based approach and Neural Network based approach for tracking a person and recognizing gestures involving arm motion. In their work they proposed illumination adaptation methods but did not consider user or hand pose adaptation. Bhuiyan et. a.l detected and tracked face and eye for human robot interaction [Bhuiyan, 2004]. But only the largest skin-like region for the probable face has been considered, which may not be true when two hands are present in the image. However, all of the above papers focus primarily on visual processing and do not maintain knowledge of different users nor consider how to deal with them.

2.2 Face detection and recognition

A first step of any face recognition or visually person identification system is to locate the face in the images. After locating the probable face, researchers use facial features (eyes, nose, nostrils, eyebrows, mouths, leaps, etc.) detection method to detect face accurately [Yang, 2000]. Face recognition or person identification compares an input face image or image features against a known face database or features databases and report match, if any. Following two subsections summarize promising past research works in the field of face detection and recognition.

2.2.1 Face detection

Face detection from a single image or an image sequences is a difficult task due to variability in pose, size, orientation, color, expression, occlusion and lighting condition. To build a fully automated system that extracts information from images of human faces, it is essential to develop and efficient algorithms to detect human faces. Visual detecting of face has been studied extensively over the last decade. Face detection researchers summarized the face detection work into four categories: template matching approaches [Sakai, 1996] [Miao, 1999] [Augusteijn, 1993] [Yuille, 1992] [Tsukamoto, 1994]; feature invariant approaches [Sirohey, 1993]; appearance-based approaches [Turk, 1991] and knowledge-based approaches [Yang, 1994] [Yang, 2002] [Kotropoulous, 1997]. Many researches also used skin color as a feature and leading remarkably face tracking as long as the lighting conditions do not varies too much [Dai, 1996], [Crowley, 1997] [Bhuiyan, 2003], [Hasanuzzaman 2004b]. Recently, several researchers combined multiple features for face localization and detection and those are more robust than single feature based approaches. Yang and Ahuja [Yang, 1998] proposed a face detection method based on color, structure and geometry. Saber and Tekalp [Saber, 1998] presented a frontal view-face localization method based on color, shape and symmetry.

2.2.2 Face recognition

During the last few years face recognition has received significant attention from the researchers [Zhao, 2003] [Chellappa, 1995]. Zhao [Zhao, 2003] et. al. have summarized the past recent research on face recognition methods with three categories: Holistic matching methods, Feature-based matching methods and Hybrid methods. One of the most widely used representations of the face recognition is eigenfaces, which are based on principal

component analysis (PCA). The eigenface algorithm uses the principal component analysis (PCA) for dimensionality reduction and to find the vectors those are best account for the distribution of face images within the entire face image spaces. Turk and Pentland [Turk, 1991] first successfully used eigenfaces for face detection and person identification or face recognition. In this method from the known face images training image dataset are prepared. The face space is defined by the "eigenfaces" which are eigenvectors generated from the training face images. Face images are projected onto the feature space (or eigenfaces) that best encodes the variation among known face images. Recognition is performed by projecting a test image onto the "facespace" (spanned by the m number of eigenfaces) and then classified the face by comparing its position (Euclidian distance) in face space with the positions of known individuals.

Independent component analysis (ICA) is similar to PCA except that the distributions of the components are designed to be non-Gaussian. The ICA separates the high-order moments of the input in addition to the second order moments utilized in PCA [Bartlett, 1998]. Face recognition system using Linear Discriminant Analysis (LDA) or Fisher Linear Discriminant Analysis (FDA) has also been very successful [Belhumeur, 1997]. In feature-based matching methods, facial features such as the eyes, lips, nose and mouth are extracted first and their locations and local statistics (geometric shape or appearance) are fed into a structural classifier [Kanade, 1977]. One of the most successful of these methods is the Elastic Bunch Graph Matching (EBGM) system [Wiskott, 1997]. Other well-known methods in these systems are Hidden Markov Model (HMM) and convolution neural network [Rowley, 1997].

2.3 Gesture recognition and gesture based interface

A gesture is a motion of the body parts or the whole body that contains information [Billinghurst, 2002]. The first step in considering gesture-based interaction with computers or robots is to understand the role of gestures in human-human communication. Gestures are varying among individuals or vary from instance to instance for a given individual. The Gesture meanings also follow one-to-many mapping or many-to-one mapping. Two approaches are commonly used to recognize gestures. One is a glove-based approach that requires wearing of cumbersome contact devices and generally carrying a load of cables that connect the devices to computers [Sturman, 1994]. Another approach is a vision-based technique that does not require wearing any contact devices, but uses a set of video cameras and computer vision techniques to interpret gestures [Pavlovic, 1997]. Although glove-based approaches provide more accurate results, they are expensive and encumbering. Computer vision techniques overcome these limitations. In general, vision-based systems are more natural than glove-based systems and are capable of hand, face and body tracking but do not provide the same accuracy in pose determination. However, for general purposes, achieving a higher-level accuracy may be less important than a real-time and inexpensive method. In addition, many gestures involve two hands, but most of the research efforts in glove-based gesture recognition use only one glove for data acquisition. In vision-based systems, we can use two hands and facial gestures at the same time.

Vision-based gesture recognition systems have three major components: image processing or extracting important clues (hand or face pose and position), tracking the gesture features (related position or motion of face and hand poses), and gesture interpretation. Vision-based gesture recognition system varies along a number of dimensions: number of cameras, speed and latency (real-time or not), structural environment (restriction on lighting conditions and

background), primary features (color, edge, regions, moments, etc.), user requirements etc. Multiple cameras can be used to overcome occlusion problems for image acquisition but this adds correspondences and integration problems. The first phase of vision-based gesture recognition task is to select a model of the gesture. The modelling of gesture depends on the intent-dent applications by the gesture. There are two different approaches for vision-based modelling of gesture: Model based approach and Appearance based approach. The Model based techniques are tried to create a 3D model of the user hand (parameters: Joint angles and palm position) [Rehg, 1994] or contour model of the hand [Shimada, 1996] [Lin, 2002] and use these for gesture recognition. The 3D models can be classified in two large groups: volumetric model and skeletal models. Volumetric models are meant to describe the 3D visual appearance of the human hands and arms. Skeletal models are related to the human hand skeleton.

Once the model is selected, an image analysis stage is used to compute the model parameters from the image features that are extracted from single or multiple video input streams. Image analysis phase includes hand localization, hand tracking, and selection of suitable image features for computing the model parameters. Two types of cues are often used for gesture or hand localization: color cues and motion cues. Color cue is useful because human skin color footprint is more distinctive from the color of the background and human cloths [Kjeldsen, 1996], [Hasanuzzaman, 2004d]. Color-based techniques are used to track objects defined by a set of colored pixels whose saturation and values (or chrominance values) are satisfied a range of thresholds. The major drawback of color-based localization methods is that skin color footprint is varied in different lighting conditions and also the human body colors. Infrared cameras are used to overcome the limitations of skin-color based segmentation method [Oka, 2002].

The motion-based segmentation is done just subtracting the images from background [Freeman, 1996]. The limitation of this method is considered the background or camera is static. Moving objects in the video stream can be detected by inter frame differences and optical flow [Cutler, 1998]. However such a system cannot detect a stationary hand or face. To overcome the individual shortcomings some researchers use fusion of color and motion cues [Azoz, 1998]. The computation of model parameters is the last step of the gesture analysis phase and it is followed by gesture recognition phase. The type of computation depends on both the model parameters and the features that were selected. In the recognition phase, parameters are classified and interpreted in the light of the accepted model or the rules specified for the gesture interpretation. Two tasks are commonly associated with the recognition process: optimal partitioning of the parameter space and implementation of the recognition procedure. The task of optimal partitioning is usually addresses through different learning-from-examples training procedures. The key concern in the implementation of the recognition procedure is computation efficiency. A recognition method usually determines confidence scores or probabilities that define how closely the image data fits each model. Gesture recognition methods are divided into two categories: static gesture or hand poster and dynamic gesture or motion gesture.

Eigenspace or PCA is also used for hand pose classification similarly its used for face detection and recognition. Moghaddam and Pentland used eigenspaces (eigenhands) and principal component analysis not only to extract features, but also as a method to estimate complete density functions for localization [Moghaddam, 1995]. In our previous research, we have used PCA for hand pose classification from three larger skin-like components that

are segmented from the real-time capture images [Hasanuzzaman, 2004d]. Triesch et. al. [Triesch, 2002] employed the elastic graph matching techniques to classify hand posters against complex backgrounds. They represented hand posters by label graphs with an underlying two-dimensional topology. Attached to the nodes are jets, which are a sort of local image description based on Gabor filters. This approach can achieve scale-invariant and user invariant recognition and does not need hand segmentation. This approach is not view-independent, because use one graph for one hand posture. The major disadvantage of this algorithm is the high computational cost.

Appearance based approaches use template images or features from the training images (images, image geometry parameters, image motion parameters, fingertip position, etc.) which use for gesture recognition [Birk, 1997]. The gestures are modeled by relating the appearance of any gesture to the appearance of the set of predefined template gestures. A different group of appearance-based model uses 2D hand image sequences as gesture templates. For each gestures number of images are used with little orientation variations [Hasanuzzaman, 2004a]. Appearance based approaches are generally computationally less expensive than model based approaches because its does not require translation time from 2D information to 3D model. Dynamic gestures recognition is accomplished using Hidden Markov Models (HMMs), Dynamic Time Warping, Bayesian networks or other patterns recognition methods that can recognize sequences over time steps. Nam et. al. [Nam, 1996] used HMM methods for recognition of space-time hand-gestures. Darrel et. al. [Darrel, 1993] used Dynamic Time Warping method, a simplification of Hidden Markov Models (HMMs) to compare the sequences of images against previously trained sequences by adjusting the length of sequences appropriately. Cutler et. al. [Cutler, 1998] used a ruled-based system for gesture recognition in which image features are extracted by optical flow. Yang [Yang, 2000] recognizes hand gestures using motion trajectories. First they extract the two-dimensional motion in an image, and motion patterns are learned from the extracted trajectories using a time delay network.

3. Frame-based knowledge-representation system for gesture-based HRI

The 'frame-based approach' is a knowledge-based problem solving approach based on the so called, 'Frame theory', first proposed by Marvin Minsky [Minsky, 1974]. A frame is a data-structure for representing a stereotyped unit of human memory including definitive and procedural knowledge. Attached to each frame there are several kinds of information about the particular object or concept it describes such as name and a set of attributes called slots. Collections of related frames are linked together into frame systems. Framed-based approach has been used successfully in many robotic applications [Ueno, 2002]. Ueno presented the concepts and methodology of knowledge modeling based on Cognitive Science for realizing the autonomous humanoid service robotics arm and hand system HARIS [Ueno, 2000]. A knowledge-based software platform called SPAK (Software Platform for Agent and Knowledge management) has developed for intelligent service robots under the internet-based distributed environment [Ampornaramveth, 2004]. SPAK has been developed to be a platform on which various software components for different robotic tasks can be integrated over a networked environment. SPAK works as a knowledge and data management system, communication channel, intelligent recognizer, intelligent scheduler, and so on. Zhang et. al. [Zhang, 2004b] have developed an Industrial Robot Arm control system using SPAK. In that system SPAK works as a communication channel and intelligent robot actions scheduler. Kiatisevi et. al. [Kiatisevi, 2004] has proposed a

distributed architecture for knowledge-based interactive robots and through SPAK they have implemented dialogue-based human-robot ('Robovie') interaction for greeting scenarios. This system employs SPAK, a frame-based knowledge engine, connecting to a group of network software agents such as 'Face recognizers', 'Gesture recognizers', 'Voice recognizers', 'Robot Controller', etc. Using information received from these agents, and based on the predefined frame knowledge hierarchy, SPAK inference engine determines the actions to be taken and submit corresponding commands to the target robot control agents.

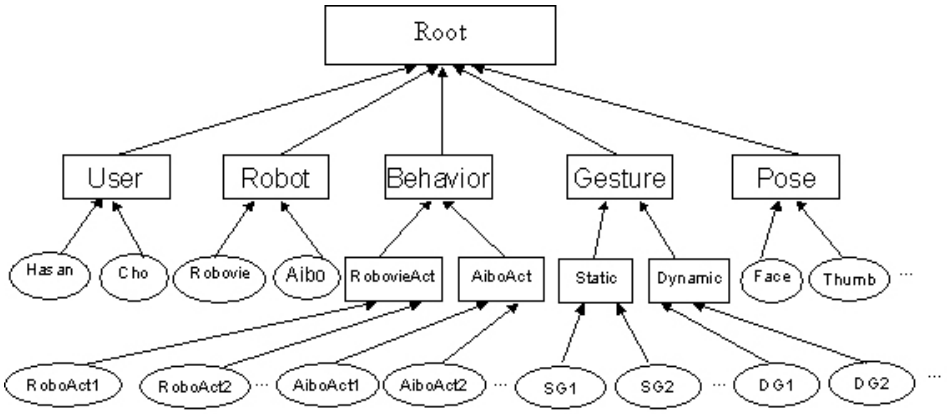


Fig. 1. Frame hierarchy for gesture-based human-robot interaction system (SG=Static Gesture, DG=Dynamic Gesture)

Figure 1 shows the frame hierarchy of the knowledge model for the gesture-based human-robot interaction system, organized by the IS_A relationship indicated by arrows connecting upper and lower frame boxes. Necessary frames are defined for the users, robots, poses, gestures and robot behaviors (actions). The user frame includes instances of all known users (instance "Hasan", "Cho", …); robot frame includes instances of all the robots ("Aibo", "Robovie",….) used in the system. The behavior frame can be sub-classed further into "AiboAct" and "RobovieAct", where "AiboAct" frame includes instances of all the predefined 'Aibo' actions and "RobovieAct" frame includes instances of all the predefined 'Robovie' actions. The gesture fame is sub-classed into "Static" and "Dynamic" for static and dynamic gestures respectively. Examples of static frame instances are, "TwoHand", "One", etc. Examples of dynamic frame instances are "YES", "NO", etc. The pose frame includes all recognizable poses such as "LEFTHAND", "RIGHTHAND", "FACE", "ONE", etc. Gesture and user frames are activated when SPAK receives information from a network agent indicating a gesture and a face has been recognized. Behavior frames are activated when the predefined conditions are met. In this model each recognizable pose is treated as an instance-frame under the class-frame "Pose". All the known poses are defined as frames. If a predefined pose is classified by vision-based pose classification module, then corresponding pose-frame will be activated. The static gestures are defined using the combination of face and hand poses. These gesture frames has three slots for the gesture components.

Each robot behaviour, includes a command or series of commands for a particular task. Each robot behavior is mapped with user and gesturer (user-gesture-action) as we proposed person-centric interpretation of gesture. In this model same gesture can be used to activate different actions of a robot for different persons even the robot is same.

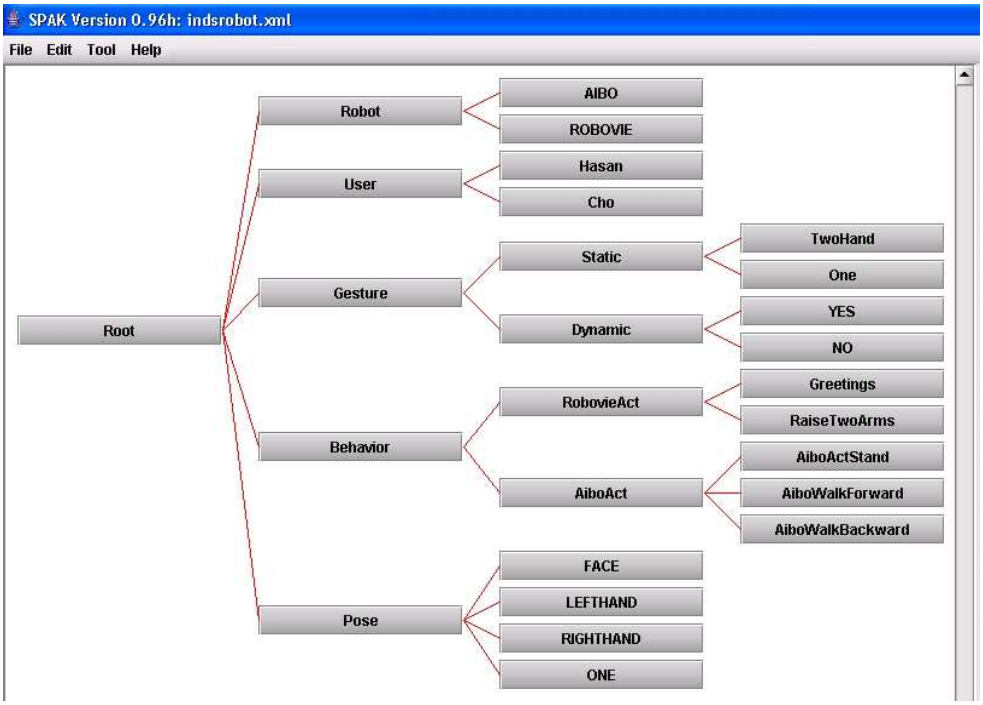


Fig. 2. Knowledge Editor showing the example gesture-based human-robot interaction

The frame system can be created in Graphical User Interface (GUI) and interpreted by the SPAK inference engine. Figure 2 shows the example knowledge Editor (part of the system) with the example of 'Robot', 'User', 'Gesture', 'Behaviour' (robot actions) and 'Pose' frames. The knowledge Editor Window displays the current frame hierarchy. Each frame is represented as click-able button. The buttons are connected with lines indicating IS_A relationships among the frames. Clicking on the frame button brings up its slot editor. Figure 3 shows an example of slot editor for the robot (Robovie) action-frame "RaiseTwoArms". The attributes for a Slot are defined by slot name, type, value, condition, argument, required, shared, etc. Figure 4 shows an example instance-frame 'Hasan' of the class-frame 'User' defined in SPAK. Figure 5 shows an example instance-frame 'Robovie' of class-frame 'Robot' defined in SPAK.

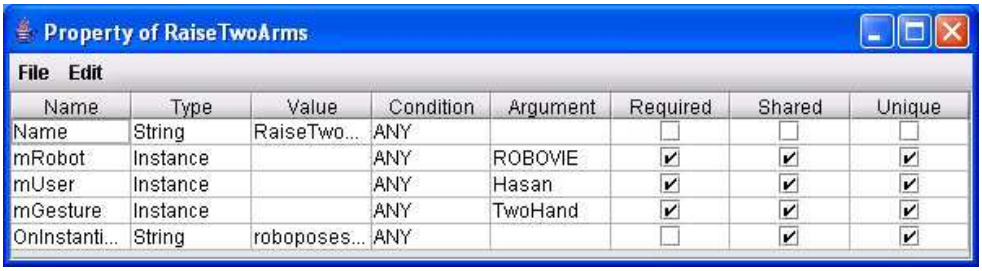
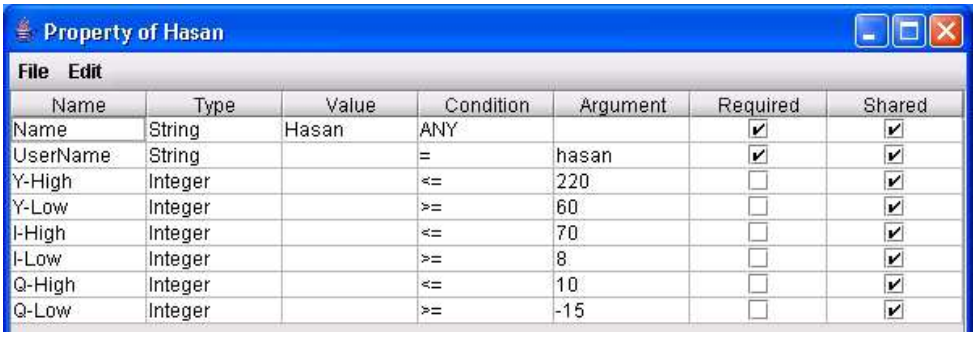


Fig. 3. Example of a robot action-frame 'RaiseTwoArms' in SPAK

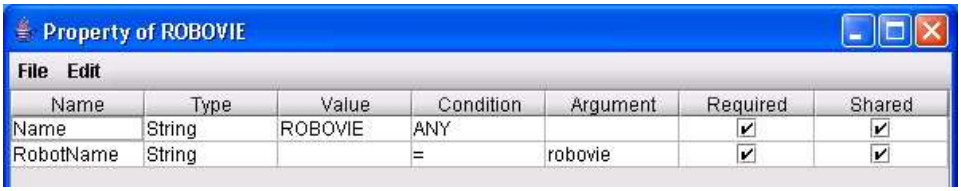


Property of Hasan

File Edit

Name	Type	Value	Condition	Argument	Required	Shared
Name	String	Hasan	ANY		✔	✔
UserName	String		=	hasan	✔	✔
Y-High	Integer		<=	220	☐	✔
Y-Low	Integer		>=	60	☐	✔
I-High	Integer		<=	70	☐	✔
I-Low	Integer		>=	8	☐	✔
Q-High	Integer		<=	10	☐	✔
Q-Low	Integer		>=	-15	☐	✔

Fig. 4. Example of instance-frame 'Hasan' of class-frame 'User' in SPAK

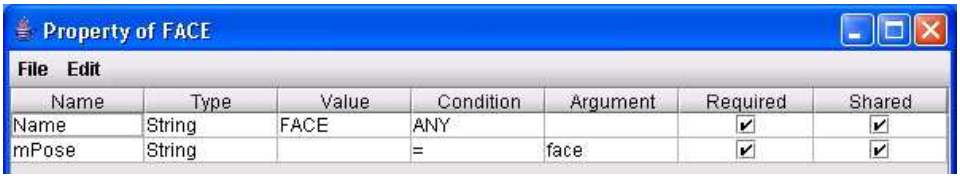


Property of ROBOVIE

File Edit

Name	Type	Value	Condition	Argument	Required	Shared
Name	String	ROBOVIE	ANY		✔	✔
RobotName	String		=	robovie	✔	✔

Fig. 5. Example of instance-frame 'Robovie' of class-frame 'Robot' in SPAK



Property of FACE

File Edit

Name	Type	Value	Condition	Argument	Required	Shared
Name	String	FACE	ANY		✔	✔
mPose	String		=	face	✔	✔

Fig. 6. Example of instance-frame 'FACE' of class-frame 'Pose' in SPAK

Property of LEFTHAND

File Edit

Name	Type	Value	Condition	Argument	Required	Shared
Name	String	LEFTHAND	ANY		✔	✔
mPose	String		=	lefthand	✔	✔

Fig. 7. Example of instance-frame 'LEFTHAND' of class-frame 'Pose' in SPAK

Image analysis module classifies the hand and face poses and identifies the user. Image analysis module, sends user name (hasan, cho, etc.) and pose names (face, lefthand, righthand, etc.) to the SPAK knowledge module. According to pose name and user name corresponding pose frame and user frame will be activated. Figure 6, Figure 7, and Figure 8 shows the instance-frames 'FACE', 'LEFTHAND', 'RIGHTHNAD' of the class-frame 'Pose' respectively. If the required combination of the pose components is found then the corresponding gesture frame will be activated. Figure 9 shows the gesture frame 'TwoHand' in SPAK. It will be activated if pose fames 'FACE', 'LEFTHAND' and 'RIGHTHNAD' are

activated. Using the received gesture and user information, SPAK processes the facts and activates the corresponding robot action frames to carry out predefined robot actions, which may include body movement and speech. The robot behaviour is user dependent and it is mapped based on user and gesture relationship (user-gesturer-robot-action) in the knowledge base. Figure 3 shows an example of 'Robovie' robot action-frame for the action "Raise Two Arms". This frame only activated if the identified user is 'Hasan', recognized gesture is "TwoHand" and the selected robot is 'Robovie'. User can add or edit the necessary knowledge frames for the users, face and hand poses, gestures and robot behaviors using SPAK knowledge Editor. The new robot behaviour frame can be included in the knowledge base according to generalization of multiple occurrences of the same gesture with user consent (first time only).

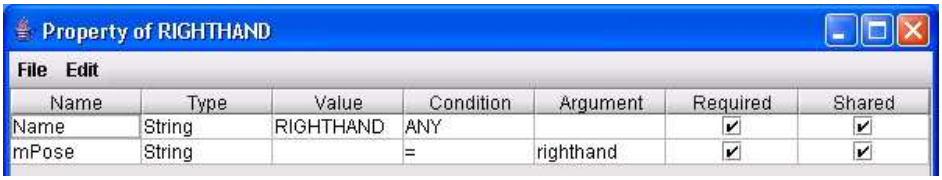


Fig. 8. Example of instance-frame 'RIGHTHAND' of class-frame 'Pose' in SPAK

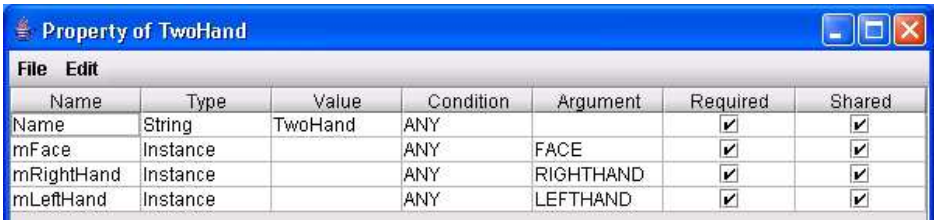


Fig. 9. Example of instance-frame 'TwoHand' of class-frame 'Gesture' in SPAK

4. Users, poses, gesture and robot behavior adaptation

This section describes new users, poses, gestures and robot behaviours adaptation methods for implementing human-robot interaction. Suppose, the robot fixed with a same room with same lighting condition, in that case only the user skin color dominates on the color-based face and hands segmentation method. It is essential for the system to cope with the different persons. The new user may not be included in the system during training phase, so the person should be included using on-line registration process. The user may want to perform new gestures that is ever been used by others person or himself. In that case the system should include the new poses with minimal user interaction. The system learns new users, new poses using multi-clustering approach with minimum user interaction. To adapt to new users and new hand poses the system must be able to perceive and extract relevant properties from the unknown faces and hand poses, find common patterns among them and formulate discrimination criteria consistent with the goals of the recognition process. This form of learning is known as clustering and it is the first steps in any recognition process where discriminating features of the objects are not know in advance [Patterson, 1990]. Subsection 4.1 describes multi-clustering based learning method.

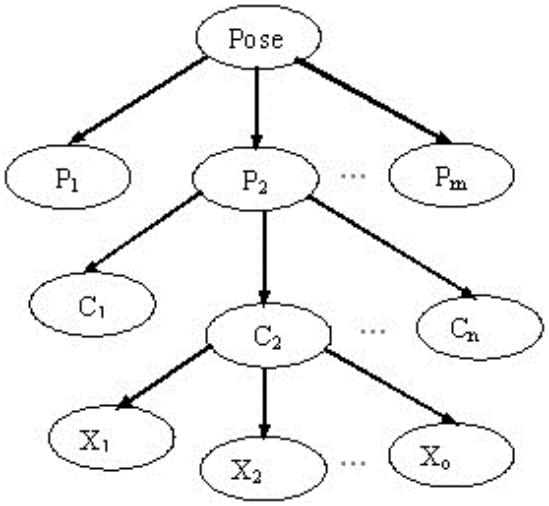


Fig. 10. Multi-cluster hierarchies for object classification and learning

4.1 Multi-clustering based learning method

Figure 10 shows the conceptual hierarchy of face and hand-pose learning using multi-cluster approach. A pose Pi may include number of clusters and each cluster Cj may include number of images (X1, X2, …, Xo) as a member of that cluster. This clustering method is described using following steps:

Step 1. Generate eigenvectors from training images that includes all the known hand poses [Turk, 1991].

Step 2. Select m-number of eigenvectors corresponding to the higher order of eigenvalues. These selected eigenvectors are regarded as principal components. The eigenvalues are sorted from high to low values.

Step 3. Read the initial cluster image database (initialize with the known cluster images) and cluster information table that's hold the starting pointer of each cluster. Project each image onto the eigenspaces and form feature vectors using equation (1) and (2).

$$\omega_i^j = (u_m)^T (T_j) \tag{1}$$

$$\Omega_j = [\omega_1^j, \omega_2^j, ..., \omega_k^j] \tag{2}$$

Where, (u_m) is the *m-th* eigenvectors, T_j is the images (60×60) in the cluster database.

Step 4. Read the unlabeled images those should be clustered or labeled.

a. Project each unlabeled image onto the eigenspaces and form feature vectors (Ω) using equation (1) and (2).

b. Calculate Euclidean distance to each image in the known (clustered) dataset using equation (3) and (4),

$$\varepsilon_j = | \, | \Omega - \Omega_j | \, | \tag{3}$$

$$\varepsilon = \arg\min\{\varepsilon_j\} \tag{4}$$

Step 5. Find the nearest class,

a. If (Ti=< ε <=Tc) then add the image in the neighbor cluster; increment the insertion parameter.

b. If (ε <Ti), then the image is recognizable and no need to include it in the cluster database.

Step 6. If the insertion rate into the known cluster is grater than zero, then update the cluster information table that's holds the starting pointer of all clusters.

Step 7. Do the step 3 to step 6 until the insertion rate (α) into the known cluster (training) data set is zero (0).

Step 8. If insertion rate is zero, then cheek the unlabeled dataset, which follows the condition (Tc < ε <=Tf). Where, Tf is the threshold that defines for discarding the image.

Step 9. If maximum number of unlabeled data (for a class)>N (predefined), then select one image (based on minimum Euclidian distance) as a member of a new cluster. Then update the cluster information table.

Step 10. Repeat from step 3 to step 9 until the number of unlabeled data less than N.

Step 11. If height of the cluster (number of member images in the cluster) is>L, then add its as a new cluster.

Step 12. After clustering poses, the user defines the association of the clusters in the knowledge base. Each pose may be associated with multiple clusters. For undefined cluster there is no association link.

4.2 Face recognition and user adaptation

We have already mentioned that the robot should be able to recognize and remember the people and learn about them [Aryananda, 2002]. If the new user comes in front of the robot eyes camera or system camera the system identifies the person as unknown person and asks for registration. The face is first detected from the cluttered background using multiple feature-based approaches [Hasanuzaman, 2007]. The detected face is filtered in order to remove noises and normalized so that it matches with the size and type of the training image [Hasanuzzaman, 2006]. The detected face is scaled to be a square image with 60x60 dimensions and converted to be a gray image. The face pattern is classified using the eigenface method [Turk, 1991], whether it belongs to known person or unknown person. The eigenvectors are calculated from the known persons face images for all face class and m-number of eigenvectors corresponding to the highest eigenvalues are chosen to form principal components for each class. The Euclidean distance is determined between the weight vectors generated from the training images and the weight vectors generated from the detected face by projecting them onto the eigenspaces. If the minimal Euclidian distance is less than the predefined threshold value then person is known, otherwise unknown. For unknown person based on judge function learning process is activated and the system learned new user using multi-clustering approach [Section 4.1]. The judge function is based on the ratio of the number of unknown face to total number of detected faces for a specific time slot. The learning function develops new cluster/clusters corresponding to a new person. The user defines the person name and skin color information in the user profile knowledge base and associates with the corresponding cluster. For known user, person-centric skin color information (Y, I, Q components) is used to reduce the computational cost.

Fig. 11. Example output of multi-clustering algorithm for learning new user

Figure 11 shows the example output of multi-clustering approach for recognizing and learning new user. For example, the system is initially trained by the face images (100 face images with five directions) of person_1. The system learns and remembers this person using five clusters (FC1, FC2, FC3, FC4, FC5) as shown in top five rows of Figure 11 and clusters information table (that's hold the starting position of each cluster and end position of last cluster) contents are [1, 12, 17, 27, 44, 52]. For example, in the case of face classification, if any face image matches with the known member between 12 and 16 then classified as face cluster_2 (FC2). If the face is classified as any of these five clusters, the person is identified as person_1. Suppose, the system is deal with new person 100 face images and it could not identify the person and activate the learning function. Then the system develops new six clusters (FC6, FC7, FC8, FC9, FC10, FC11) and updates the cluster information table ([1, 12, 17, 27, 44, 53, 68, 82, 86, 96, 106, 116]) as shown in Figure 11 (rows 6 to rows 11). The new user is registered as person_2 and the associations with the clusters are defined. If any detected face images is classified as known cluster then corresponding person is identified.

4.3 Hand pose classification and adaptation

For machine it is difficult to understand the new poses without prior knowledge. It is essential to learn new poses based on specific judge function or predefined knowledge. The judge function determines the user intention, i.e., intention to create new gesture. The judge function is based on the ratio of the number of unknown hand poses to the total number hand poses for a specific time slot. For example, the user shows same hand pose

continuously for 10 image frames that are unknown to the system that means he/she wants to use it as a gesture. In this proposed system the hand poses are classified using multi-cluster based learning method. For unknown pose, based on judge function learning function is activated and the system learns new pose. The learning function develops new cluster/clusters corresponding to new pose. The user defines the pose name in the knowledge base and associates with the corresponding cluster. If the pose is identified then corresponding frame will be activated. Figure 12 presents example output of learning new pose using multi-cluster approach. The system is first trained by pose 'ONE' and form one cluster for that pose. Then the system is trained by pose 'FIST' where formed another two clusters because the user uses two hands for that pose. Similarly other clusters are formed corresponding to new poses.

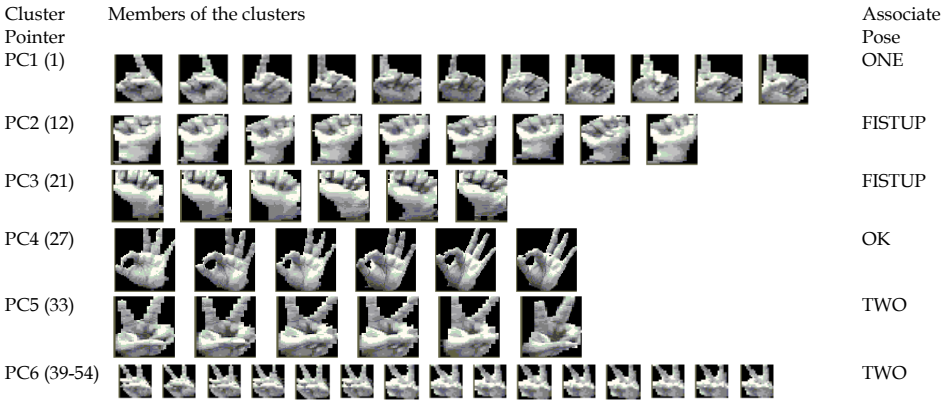


Fig. 12. Example output of multi-clustering approach for learning new pose

4.4 Gesture recognition and adaptation

The recognition of gesture is carried out in two phases. In the first phase, face and hand poses are classified from captured image frame using the method described in previous section. Then combinations of poses are analyzed to identify the occurrence of gesture. For example, if left hand palm, right hand palm and one face are present in the input image then it recognizes as "TwoHand" gesture [Figure 19(a)] and corresponding gesture frame will be activated. Interpretation of identified gesture is user-dependent since the meaning of the gesture may differ from person to person based on their culture. For example, when user'Hasan' comes in front of 'Robovie' eyes, 'Robovie' recognizes the person and says "Hi Hasan! How are you?", then 'Hasan' raises his 'Thumb up' and "Robovie" replies to 'Hasan'"Oh! You are not fine today". In the similar situation, for another user 'Cho', 'Robovie' says, "Hi, You are fine today?". That means 'Robovie' can understand the person-centric meaning of gesture. To accommodate different user's desires, our person-centric gesture interpretation is implemented using frame-based knowledge representation approach. The user predefines these frames into the knowledge base with necessary attributes (gesture components, gesture name) for all predefined gestures. Our current system recognizes 11 static gestures. These are: 'TwoHand' (raise left hand and right hand palms), 'LeftHand' (raise left hand palm), 'RightHand' (raise right hand palm), 'One' (raise

index finger), 'Two' (form V sign using index and middle fingers), 'Three' (raise index, middle and ring fingers), 'ThumbUp' (thumb up), 'Ok' (make circle using thumb and index finger), 'FistUp' (fist up), 'PointLeft' (point left by index finger), 'PointRight' (point right by index finger).
It is possible to recognize more gestures including new poses and new rules for the gesture using this system. New poses can be included in the training image database using the interactive learning method and corresponding frame can be defined in the knowledge base to interpret the gesture. To teach the robot a new poses, the user should perform the poses several times (example 10 image frame times.). Then the learning method detects it as a new pose and creates cluster/clusters for that pose. Sequentially, it updates the knowledge base for the cluster information.

4.5 Robot behaviours adaptation

Robot behaviours or actions can be programmed or learned through experience. But it is difficult to perceive human or user intention to acts robot with his/her gestures. This system has proposed the experience based and user-interactive robot behaviour learning or adaptation method. In this method the history of the similar gesture-action map is stored in the knowledge base. According to maximum user desires action will be select for the gesture and ask for the user acknowledgement. If user consents or uses "YES" gesture (or types "Yes") the corresponding frame will be store permanently. For Example scenario:
Person_n: comes in front of Robovie.
Robovie: cannot recognize, and asks, "Who are you?"
Person_n: types "Person_n" (or says "Person_n").
Robovie: Says, "Hello Person_n, do you want to play with me?"
Person_n: shows "OK" hand gesture (make circle using thumb and index fingure).
Robovie: asks "Do you mean Yes"?
Person_n: again shows "OK" gesture.
Robovie: add "Ok ="Yes, for Person_n" into his knowledge base.
Suppose there are two users already use 'OK' gesture to mean Yes, so Robovie adds "OK=Yes for everybody" into his knowledge base.

5. Experimental results and discussions

This system uses a standard video camera and 'Robovie' eye's camera for data acquisition. Each captured image is digitized into a matrix of 320×240 pixels with 24-bit color. User and hand pose adaptation method is verified using real-time captured images as well as static images. The algorithm has also been tested with a real world human-robot interaction system using a humanoid robot, 'Robovie' developed by ATR .

5.1 Results of user recognition and adaptation

Seven individuals were asked to act for the predefined face poses in front of the camera and all the sequence of face images were saved as individual image frame. All the training and test images are 60x60 pixels gray face images. The adaptation algorithm is tested for 7 persons frontal face or normal face (NF) images, and five directional face images (normal face, left directed face, right directed face, up directed face, down directed face). Figure 13

shows the sample result of the user adaptation method for normal faces. In the first step, the system is trained using 60 face images of three persons and developed three clusters (top 3 rows of Figure 13) corresponding to three persons. The cluster information table contents are [1, 11, 23, 29]. For example, in this situation if any input face image matches with the known face image member between 1 and 10 then the person is identified as person 1.

Fig. 13. Sample outputs of the clustering method for frontal faces

In the second step, 20-face image sequences of another person are fed to the system as input. The minimum Euclidian distances (ED) from three known persons face images are shown using upper line graph (B_adap) in Figure 14. The system identifies these faces as unknown person based on predefined threshold value for the Euclidian distance and activates the user learning function. The user learning function developed new cluster (4th row of Figure 13) and updated the cluster information table as [1, 11, 23, 30, 37]. After adaptation, the minimum Euclidian distance distribution line (A_adap) in Figure 7.21 shows that, for 8 images, minimum ED is zero and those are included in the new cluster so that the system can recognize the person. This method is tested for 7 persons including 2 females, and as a result of learning, 7 clusters with different length (number of images per cluster) for different persons (as shown in Figure 13) were formed.

The users adaptation method is also tested for 700 five directional face images of 7 persons (sample output in Figure 11). Figure 15 shows the distribution of 41 clusters for the 700 face images of 7 persons. In the first step, the system is trained using 100 face images of person_1 and it formed 5 clusters based on 5-directional faces. At this time, the contents of the cluster information table (that holds staring pointer of each cluster in the training database) are [1, 12, 17, 27, 44, 52]. After learning person_2, the cluster information table contents are [1, 12, 17, 27, 44, 53, 68, 82, 86, 96, 106, 116]. Similarly, other persons are adapted. Figure 16 shows the example of errors in clustering process. In the cluster 26 up directed faces of person_6 and frontal face person_5 are overlapped and treated as one cluster (Figure 16 (a)). In the case of cluster 31, up directed faces of person_5 and normal (frontal) faces of person_6 are overlapped and grouped in the same cluster (Figure 16 (b)). This problem can be solved using narrow threshold, but in that case the number of iteration as well as discard rates of the images classification method will be increased.

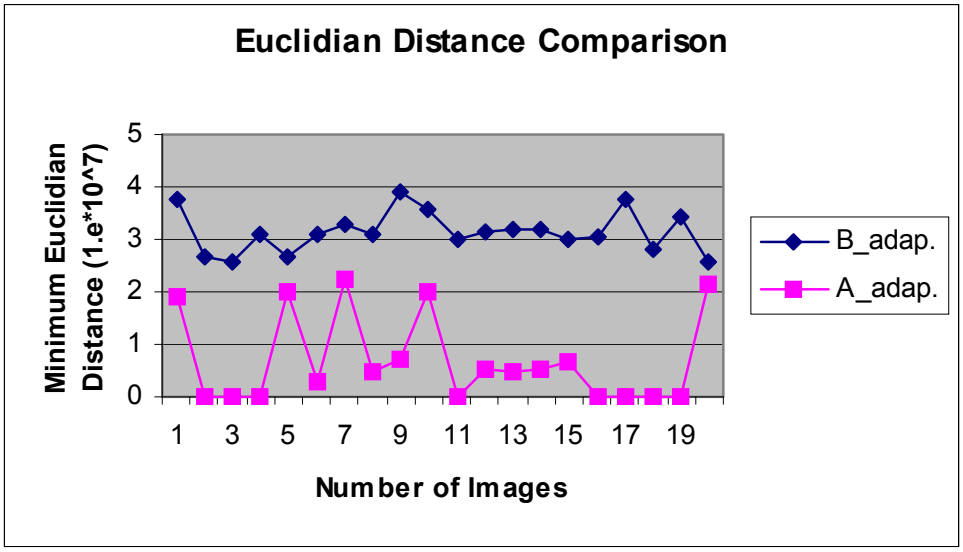


Fig. 14. Euclidian distances distribution of 20 frontal faces (before and after adaptation)

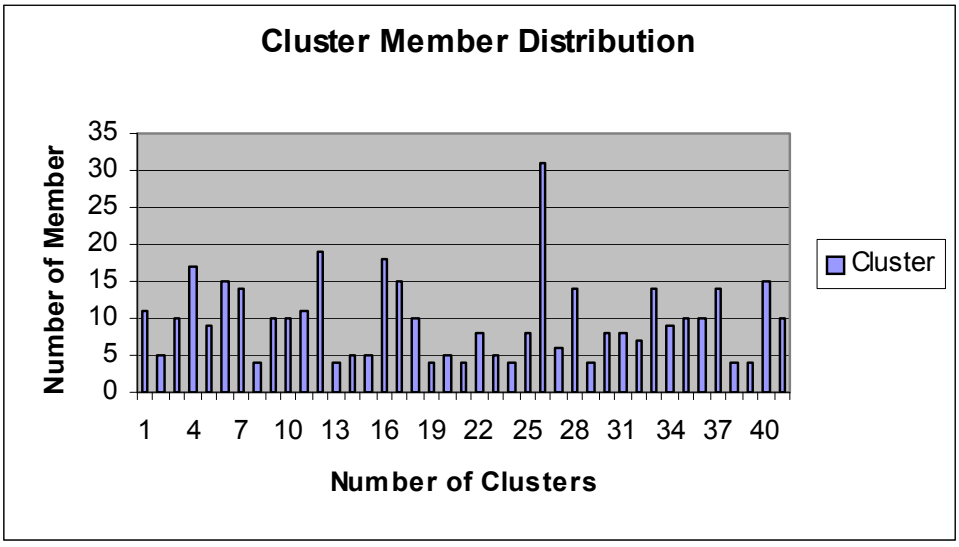


Fig. 15. Cluster member distributions for five directed face poses

(a) Cluster 26: Person_5 frontal face and person _6 up directed face overlapping (231-261)

(b) Cluster 31: Person_5 up face and person_6 (from right 2 image) frontal face overlapping

Fig. 16. Example of errors in clustering process

5.2 Results of pose classification and adaptation

The system uses 10 hand poses of 7 persons for evaluating pose classification and adaptation method. All the training and test images are 60x60 pixels gray images. This system is first trained using 200 images of 10 poses of person_1 (20 images of each pose). It automatically clusters the images into 13 clusters. Figure 12 shows the sample outputs of hand poses learning method for person_1. If the user uses two hands to make the same pose then it forms two different clusters for the same pose. Different clusters can also be formed for the variation of orientation even the pose is same.

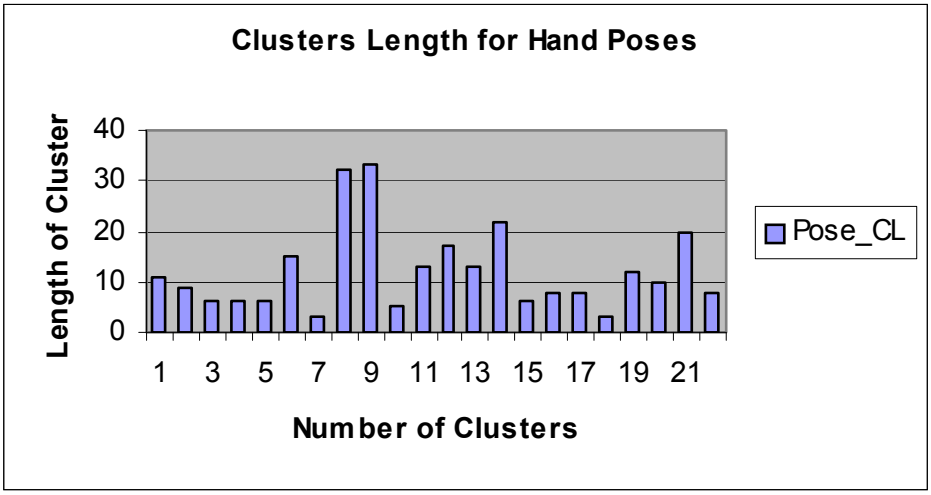


Fig. 17. Cluster member distributions for hand poses

If the person is change, then it may form different clusters for the same hand poses (gestures) due to the variation of hand shape and color. After trained with 10 hand poses of person_1, 200 images of 10 hand poses of person_2 are feed to the system. The system developed 9 more clusters for the person_2 corresponding to 8 hand poses. For, the 'LEFTHAND' and 'RIGHTHAND' palm it did not develop new clusters, rather inserted new members in those clusters. Table 1 shows the 22 clusters developed for 10 hand poses of 2 persons and their associations with hand poses. Figure 17 shows the distributions of the clusters of 10 hand poses for two persons.

Pose Name	Associated Clusters
ONE (Raise Index Finger)	PC_1, PC_{15}
FIST (Fist Up)	PC_2, PC_3, PC_{19}
OK (Make circle using thumb and index fingers)	PC_4, PC_{20}
TWO (V sign using index and middle fingers)	PC_5, PC_6, PC_{16}
THREE (Raise index, middle and ring fingers)	PC_7, PC_{17}, PC_{18}
LEFTHAND (Left hand palm)	PC_8
RIGHTHAND (Right hand palm)	PC_9
THUMB (Thumb Up)	PC_{10}, PC_{11}, PC_{14}
POINTL (Point Left)	PC_{12}, PC_{21}
POINTL (Point Right)	PC_{13}, PC_{22}

Table 1. List of hand poses and associated clusters

Input Images (ASL Char)	Number of Image	Correct Recognition		Accuracy (%)	
		Before_Adap	*After_Adap*	*B_Adap*	*A_Adap*
A	120	81	119	67.50	99.16
B	123	82	109	66.66	88.61
C	110	73	103	66.36	93.63
D	120	78	106	65	88.33
E	127	80	96	62.99	75.59
F	120	81	114	67.50	95
G	120	118	120	98.33	100
I	100	56	100	56	100
K	120	107	116	89.16	96.66
L	120	100	119	83.33	99.16
P	120	79	119	65.83	99.16
V	120	75	86	62.50	71.66
W	120	74	101	61.66	84.16
Y	120	85	98	70.83	81.66

Table 2. Comparison of pose classification accuracy (before and after adaptation) for 14 ASL Characters

In this study we have also compared pose classification accuracy (%) using two methods: one is multi-cluster based approach without adaptation, the other is multi-cluster based approach with adaptation. In this experiment we have used total 840 training images, (20 images of each pose of each person and 1660 test images of 14 ASL characters [ASL, 2004] of three persons. Table 2 presents the comparisons of two methods (B_Adap=before adaptation and A_Adap=after adaptation). This table shows that the accuracy of pose classification method which includes the adaptation or learning approach is better, because the learning function increments the clusters members or forms new clusters if necessary to classify the new images.

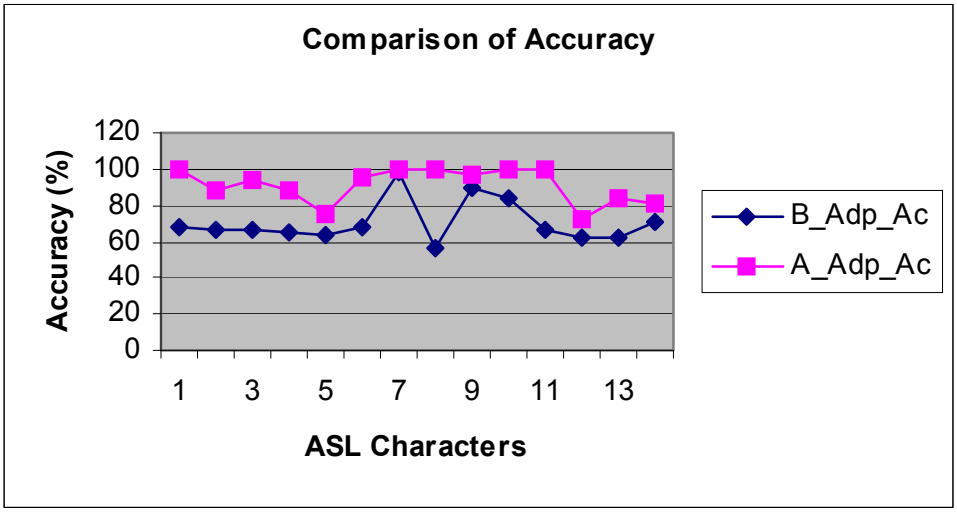


Fig. 18. Comparison of pose classification accuracy before after adaptation

Figure 18 depicts the graphical representations of 14-ASL characters classification accuracy using adaptation method (after adaptation) and without adaptation method (before adaptation). The comparison curves shows that if we include adaptation method then pose classification performance will be better, but needs user interaction that is bottleneck of this method.

6. Implementation of human-robot interaction

The real-time gesture based human-robot interaction is implemented as an application of this system. This approach has been implemented on a humanoid robot, name"Robovie". Since the same gestures can mean different tasks for different persons, we need to maintain the gesture with person-to-task knowledge. The robot and the gesture recognition PC are connected to SPAK knowledge server. From the image analysis and recognition PC, person identity and pose names are sent to the SPAK for decisions making and the robot activation. According to gesture and user identity, the knowledge module generates executable codes for robot actions. The robot then follows speech and body action commands. This method has been implemented for the following scenario:

User: "Person_1" comes in front of Robovie eyes camera and robot recognizes the user as "Person_1".

Robot: "Hi Person_1, How are you?" (speech)

Person_1: uses the gesture "Ok"

Robot: " Oh Good! Do you want to play now?" (speech)

Person_1: uses the gesture "YES"

Robot: "Oh Thanks" (speech)

Person_1: uses the gesture "TwoHand"

Robot: imitates user's gesture "Raise Two Arms" as shown in Figure 19.

Person_1: uses the gesture "FistUp" (stop the interaction)
Robot: Bye-bye (speech).
User: "Person_2" comes in front of Robovie eyes camera, robot detects the face as unknown,
Robot: "Hi, What is your Name?" (speech)
Person_2: Types his name "Person_2"
Robot: " Oh, Good! Do you want to play now?" (speech)
Person_2: uses the gesture "OK"
Robot: "Thanks!" (speech)
Person_2: uses the gesture "LeftHand"
Robot: imitate user's gesture ("Raise Left Arm")
Person_2: uses the gesture "RightHand"
Robot: imitate user's gesture ("Raise Right Arm")
Person_2: uses the gesture "Three"
Robot: This is three (speech)
Person_2: uses the gesture "TwoHand"
Robot: Bye-bye (speech)
The above scenario shows that the same gesture can be used to represent different meanings and several gestures can be used to denote the same meaning for different persons. A user can design new actions according to his/her desires using 'Robovie' and can design corresponding knowledge frames using SPAK to implement their desired actions.

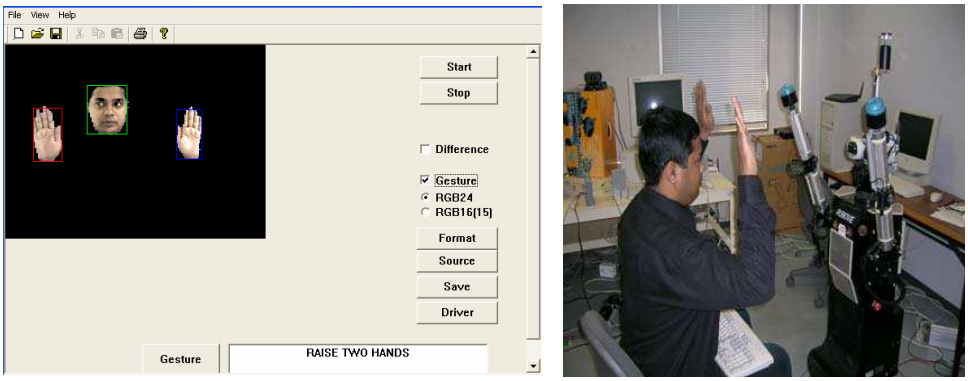


Fig. 19. Example human-robot (Robovie) interaction

7. Conclusion

This chapter describes users, gestures and robot behaviour adaptation method for human robot interaction by integrating computer vision and a knowledge-based software platform. In this method the user defines frames for users, poses, gestures, robots and robot behaviours. This chapter presents a multi-cluster based interactive learning approach for adapting new users and hand poses. However, if a large number of users use a large number of hand poses it is impossible to run this systems in real time. To overcome this

problem, in future we should maintain person-specific subspaces (individual PCA for each person of all hand poses) for pose classification and learning.
In this chapter we have also described how the system can adapt with new gestures and new robot behaviours using multiple occurrences of the same gesture with user interaction. The future aim is to make the system more robust, dynamically adaptable to new users and new gestures for interaction with different robots such as Aibo, Robovie, Scout, etc. Ultimate goal of this research is to establish a human-robot symbiotic society so that they can share their resources and work cooperatively with human beings.

8. Acknowledgment

I would like to express deep sense of gratitude and thanks to Dr. Haruki Ueno, Professor, Department of Informatics, National Institute of Informatics, Tokyo, Japan, for his sagacious guidance, encouragement and every possible help throughout this research work. I am grateful to Dr. Y. Shirai, Professor, Department of Human and Computer Intelligence, School of Information Science and Engineering, Ritsumeikan University, for his ingenious inspiration, suggestions and care in the whole research period. I would also like to express my sincere thanks to Professor H. Gotoda, Department of Informatics, National Institute of Informatics,Tokyo, Japan, for his valuable suggestions throughout the research. I must also thank to Dr. T. Zhang and Dr. V. Ampornaramveth for their assistance and suggestions during this research work.

9. References

[Ampornaramveth, 2004] V. Ampornaramveth, P. Kiatisevi, H. Ueno, "SPAK: Software Platform for Agents and Knowledge Systems in Symbiotic Robots", IEICE Transactions on Information and systems, Vol.E86-D, No.3, pp 1-10, 2004.

[Ampornaramveth, 2001] V. Ampornaramveth, H. Ueno, "Software Platform for Symbiotic Operations of Human and Networked Robots", NII Journal, Vol.3, No.1, pp 73-81, 2001.

[Aryananda, 2002] L. Aryananda, "Recognizing and Remembering Individuals: Online and Unsupervised Face Recognition for Humanoid Robot" in Proceeding of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002), Vol. 2, pp. 1202-1207, 2002.

[ASL, 2004] "American Sign Language Browser" http://commtechlab.msu.edu/sites/aslweb/browser.htm, visited on April 2004.

[Augusteijn, 1993] M. F. Augusteijn, and T.L. Skujca, "Identification of Human Faces Through Texture-Based Feature Recognition and Neural Network Technology", in Proceeding of IEEE conference on Neural Networks, pp.392-398, 1993.

[Azoz, 1998] Y. Azoz, L. Devi, and R. Sharma, "Reliable Tracking of Human Arm Dynamics by Multiple Cue Integration and Constraint Fusion", in Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'98), pp. 905-910, 1998.

[Bartlett, 1998] M. S Bartlett, H. M. Lades, and, T. Sejnowski, "Independent Component Representation for Face Recognition" in Proceedings of Symposium on Electronic Imaging (SPEI): Science and Technology, pp. 528-539, 1998.

[Belhumeur, 1997] P.N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, " Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection" IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 19, pp. 711-720, 1997.

[Bhuiyan, 2004] M. A. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno, "On Tracking of Eye For Human-Robot Interface", International Journal of Robotics and Automation, Vol. 19, No. 1, pp. 42-54, 2004.

[Bhuiyan, 2003] M. A. Bhuiyan, V. Ampornaramveth, S. Muto, H. Ueno, "Face Detection and Facial Feature Localization for Human-machine Interface", NII Journal, Vol.5, No. 1, pp. 25-39, 2003.

[Birk, 1997] H. Birk, T. B. Moeslund, and C. B. Madsen, "Real-time Recognition of Hand Alphabet Gesture Using Principal Component Analysis", in Proceeding of 10th Scandinavian Conference on Image Analysis, Finland, 1997.

[Chellappa, 1995] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of faces: A survey", in Proceeding of IEEE, Vol. 83, No. 5, pp. 705-740, 1995.

[Crowley, 1997] J. L. Crowley and F. Berard, "Multi Modal Tracking of Faces for Video Communications", In proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97), pp. 640-645, 1997.

[Cutler, 1998] R. Cutler, M. Turk, "View-based Interpretation of Real-time Optical Flow for Gesture Recognition", in Proceedings of 3rd International Conference on Automatic Face and Gesture Recognition (AFGR'98), pp. 416-421, 1998.

[Darrel, 1993] T. Darrel and A. Pentland, "Space-time Gestures", in Proceedings of IEEE International Conference on Computer Vision and Pattern recognition (CVPR'93), pp. 335-340, 1993.

[Dai, 1996] Y. Dai and Y. Nakano, "Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene", Pattern Recognition, Vol. 29, No. 6, pp.1007-1017, 1996.

[Endres, 1998] H. Endres, W. Feiten, and G. Lawitzky, "Field Test of a Navigation System: Autonomous Cleaning in Supermarkets", in Proceeding of IEEE International Conference on Robotics & Automation (ICRA '98), 1998.

[Festival, 1999] "The Festival Speech Synthesis System developed by CSTR", University of Edinburgh, http://www.cstr.ed.ac.uk/project/festival

[Fong, 2003] T. Fong, I. Nourbakhsh and K. Dautenhahn, "A Survey of Socially Interactive Robots", Robotics and Autonomous System, Vol. 42(3-4), pp.143-166, 2003.

[Freeman, 1996] W.T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma, "Computer Vision for Computer Games", in Proceedings of International Conference on Automatic Face and Gesture Recognition (AFGR'96), pp. 100-105, 1996.

[Hasanuzzaman, 2007a] Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Gotoda, Y. Shirai, H. Ueno, "Adaptive Visual Gesture Recognition for Human-Robot Interaction Using Knowledge-based Software Platform", International Journal of Robotics and Autonomous Systems (RAS), Elsevier, Vol. 55(8), pp. 643-657, 2007.

[Hasanuzzaman, 2007b] Md. Hasanuzzaman, S. M. Tareeq, Tao Zhang, V. Ampornaramveth, H. Gotoda, Y. Shirai, H. Ueno, "Adaptive Visual Gesture Recognition for Human-Robot Interaction", Malaysian Journal of Computer Science, ISSN-0127-9084, Vol. 20(1), pp. 23-34, 2007.

[Hasanuzzaman, 2006] Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, and H. Ueno, "Gesture-Based Human-Robot Interaction Using a Knowledge-Based Software Platform", International Journal of Industrial Robot, Vol. 33 (1), pp. 37-49, 2006.

[Hasanuzzaman, 2004a] M. Hasanuzzaman, V. Ampornaramveth, T. Zhang, M. A. Bhuiyan, Y. Shirai, H. Ueno, "Real-time Vision-based Gesture Recognition for Human-Robot Interaction", in Proceeding of IEEE International Conference on Robotics and Biomimetics (ROBIO'2004), China, pp. 379-384, 2004.

[Hasanuzzaman, 2004b] M. Hasanuzzaman, T. Zhang, V. Ampornaramveth, M.A. Bhuiyan, Y. Shirai, H. Ueno, "Gesture Recognition for Human-Robot Interaction Through a Knowledge Based Software Platform", in Proceeding of IEEE International Conference on Image Analysis and Recognition (ICIAR 2004), LNCS 3211 (Springer-Verlag Berlin Heidelberg), Vol. 1, pp. 5300-537, Portugal, 2004.

[Hasanuzzaman, 2004c] M. Hasanuzzaman, T. Zhang, V. Ampornaramveth, P. Kiatisevi, Y. Shirai, H. Ueno, "Gesture-based Human-Robot Interaction Using a Frame-based Software Platform", in Proceeding of IEEE International Conference on Systems Man and Cybernetics (IEEE SMC'2004), Netherland, 2004.

[Kanda, 2002] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and Evaluation of an Interactive Humanoid Robot: Robovie", in Proceeding of IEEE International Conference on Robotics and Automation (ICRA 2002), pp. 1848-1855, 2002.

[Kanade, 1977] T. Kanade, "Computer Recognition of Human Faces", Brikhauser Verlag, Basel and Stuttgart, ISR-47, pp. 1-106, 1977.

[Kiatisevi, 2004] P. Kiatisevi, V. Ampornaramveth and H. Ueno, "A Ditributed Architecture for Knowledge-Based Interactive Robots", in Proceeding of 2nd International Conference on Information Technology for Application (ICITA' 2004), pp. 256-261, 2004.

[King, 1990] S. King and C. Weirman, "Helpmate Autonomous Mobile Robot Navigation System", in Proceeding of SPIE Conference on Mobile Robots, pp. 190-198,1990.

[Kjeldsen, 1996] R. Kjeldsen, and K. Kender, "Finding Skin in Color Images", in Proceedings of 2nd International Conference on Automatic Face and Gesture Recognition (AFGR'96), pp. 312-317, 1996.

[Kortenkamp, 1996] D. Kortenkamp, E. Hubber, and P. Bonasso, "Recognizing and Interpreting Gestures on a Mobile robot", in Proceeding of AAAI'96, pp. 915-921, 1996.

[Kotropoulos, 1997] C. Kotropoulos and I. Pitas, "Rule-based Face Detection in Frontal Views", in Proceeding of International Conference on Acoustics, Speech and Signal Processing, Vol. 4, pp. 2537-2540, 1997.

[Lin, 2002] J. Lin, Y. Wu, and T. S Huang, "Capturing Human Hand Motion in Image Sequences", in Proceeding of Workshop on Motion and Video Computing, Orlando, Florida, December, 2002.

[Miao, 1999] J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen, "A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Centre Template", Pattern Recognition, Vol. 32, No. 7, pp. 1237-1248, 1999.

[Minsky, 1974] M. Minsky "A Framework for Representing Knowledge", MIT-AI Laboratory Memo 306, 1974.

[Moghaddam, 1995] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection", in Proceeding of 5th International Conference on Computer Vision, pp. 786-793, 1995.

[Nam, 1996] Y. Nam and K. Y. Wohn, "Recognition of Space-Time Hand-Gestures Using Hidden Markov Model", in Proceedings of ACM Symposium on Virtual Reality Software and Technology, pp. 51-58, 1996.

[Oka, 2002] K. Oka, Y. Sato, and H. Koike, "Real-Time Tracking of Multiple Finger-trips and Gesture Recognition for Augmented Desk Interface Systems", in Proceeding of International Conference in Automatic Face and Gesture Recognition (AFGR'02), pp. 423-428, Washington D.C, USA, 2002.

[Patterson, 1990] D. W. Patterson, "Introduction to Artificial Intelligence and Expert Systems", Prentice-Hall Inc., Englewood Cliffs, N.J, USA, 1990.

[Pavlovic, 1997] V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 19, No. 7, pp. 677-695, 1997.

[Perzanowski, 2001] D. Perzanowski, A. C. Schultz, W. Adams, A. Marsh, and M. Bugajska, "Building a Multimodal Human-Robot Interface", IEEE Intelligent Systems, Vol. 16(1), pp. 16-21, 2001.

[Pineau, 2003] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, S. Thrun, "Towards Robotic Assistants in Nursing Homes: Challenges and Results", Robotics and Autonomous Systems, Vol. 42, pp. 271-281, 2003.

[Rehg, 1994] J. M. Rehg and T. Kanade, "Digiteyes: Vision-based Hand Tracking for Human-Computer Interaction", in Proceeding of Workshop on Motion of Non-Rigid and Articulated Bodies, pp. 16-94, 1994.

[Rowley, 1998] H. A. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection" IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 23, No. 1, pp. 23-38, 1998.

[Saber, 1998] E. Saber and A. M. Tekalp, "Frontal-view Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions", Pattern Recognition Letters, Vol. 17(8) pp.669-680, 1998.

[Sakai, 1996] T. Sakai. M. Nagao and S. Fujibayashi, "Line Extraction and Pattern Detection in a Photograph", Pattern Recognition, Vol. 1, pp.233-248, 1996.

[Severinson-Eklundh, 2003] K. Severinson-Eklundh, A. Green, H. Huttenrauch, "Social and Collaborative Aspects of Interaction with a Service Robot", Robotics and Autonomous Systems, Vol. 42, pp.223-234, 2003.

[Shimada, 1996] N. Shimada, and Y. Shirai, "3-D Hand Pose Estimation and Shape Model Refinement from a Monocular Image Sequence", in Proceedings of VSMM'96 in GIFU, pp.23-428, 1996.

[Siegwart, 2003] R. Siegwart et. al., "Robox at Expo.02: A large-scale Installation of Personal Robots", Robotics and Autonomous Systems, Vol. 42, pp. 203-222, 2003.

[Sirohey, 1993] S. A. Sirohey, "Human Face Segmentation and Identification", *Technical Report* CS-TR-3176,University of Maryland, pp. 1-33, 1993.

[Sturman, 1994] D.J. Sturman and D. Zetler, "A Survey of Glove-Based Input" IEEE Computer Graphics and Applications, Vol. 14, pp-30-39, 1994.

[Tsukamoto, 1994] A. Tsukamoto, C.W. Lee, and S. Tsuji, "Detection and Pose Estimation of Human Face with Synthesized Image Models," in Proceeding of International Conference of Pattern Recognition, pp. 754-757,1994.

[Torras, 1995] C. Torras, "Robot Adaptivity", Robotics and Automation Systems, Vol. 15, pp.11-23, 1995.

[Torrance, 1994] M. C. Torrance, "Natural Communication with Robots" Master's thesis, MIT, Department of Electrical Engineering and Computer Science, Cambridge, MA, January 1994.

[Triesch, 2002] J. Triesch and C. V. Malsburg, "Classification of Hand Postures Against Complex Backgrounds Using Elastic Graph Matching", Image and Vision Computing, Vol. 20, pp. 937–943, 2002.

[Turk, 1991] M. Turk and A. Pentland, "Eigenface for Recognition" Journal of Cognitive Neuroscience, Vol. 3, No.1, pp. 71-86, 1991.

[Ueno, 2002] H. Ueno, "A Knowledge-Based Information Modeling for Autonomous Humanoid Service Robot", IEICE Transactions on Information & Systems, Vol. E85-D, No. 4, pp. 657-665, 2002.

[Ueno, 2000] H. Ueno, "A Cognitive Science-Based Knowledge Modeling for Autonomous Humanoid Service Robot-Towards a Human-Robot Symbiosis", in Proceeding of 10th European-Japanese Conference on Modeling and Knowledge Base, pp. 82-95, 2000.

[Waldherr, 2000] S. Waldherr, R. Romero, S. Thrun, "A Gesture Based Interface for Human-Robot Interaction", Journal of Autonomous Robots, Kluwer Academic Publishers, pp. 151-173, 2000.

[Weimer, 1989] D. Weimer and S. K. Ganapathy, "A Synthetic Virtual Environment with Hand Gesturing and Voice Input", in Proceedings of ACM CHI'89 Human Factors in Computing Systems, pp. 235-240, 1989.

[Wiskott, 1997] L. Wiskott, J. M. Fellous, N. Kruger, and C. V. Malsburg, "Face Recognition by Elastic Bunch Graph Matching", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 19, No.7, pp. 775-779, 1997.

[Yang, 2002] M. H. Yang, D. J. Kriegman and N. Ahuja, "Detection Faces in Images: A survey", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 24, No. 1, pp. 34-58, 2002.

[Yang, 2000] M. H. Yang, "Hand Gesture Recognition and Face Detection in Images", Ph.D Thesis, University of Illinois, Urbana-Champaign, 2000.

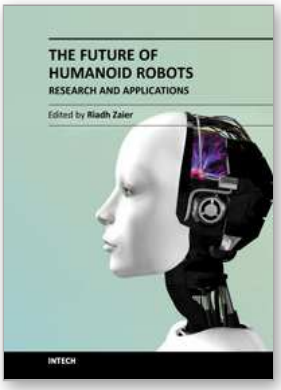
[Yang, 1998] J. Yang, R. Stiefelhagen, U. Meier and A. Waibel, "Visual Tracking for Multimodal Human Computer Interaction", in Proceedings of ACM CHI'98 Human Factors in Computing Systems, pp. 140-147, 1998.

[Yang, 1994] G. Yang and T. S. Huang, "Human Face Detection in Complex Background", Pattern Recognition, Vol. 27, No.1, pp.53-63, 1994.

[Yuille, 1992] A. Yuille, P. Hallinan and D. Cohen, "Feature Extraction from Faces Using Deformable Templates, International Journal of Computer Vision, Vol. 8, No. 2, pp 99-111, 1992.

[Zhang, 2004b] T. Zhang, M. Hasanuzzaman, V. Ampornaramveth, P. Kiatisevi, H. Ueno, "Human-Robot Interaction Control for Industrial Robot Arm Through Software Platform for Agents and Knowledge Management", in Proceedings of IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC 2004), Netherlands, pp. 2865-2870, 2004.

[Zhao, 2003] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey" ACM Computing Surveys, Vol. 35, No. 4, pp. 399-458, 2003.



The Future of Humanoid Robots - Research and Applications

Edited by Dr. Riadh Zaier

ISBN 978-953-307-951-6
Hard cover, 300 pages
Publisher InTech
Published online 20, January, 2012
Published in print edition January, 2012

This book provides state of the art scientific and engineering research findings and developments in the field of humanoid robotics and its applications. It is expected that humanoids will change the way we interact with machines, and will have the ability to blend perfectly into an environment already designed for humans. The book contains chapters that aim to discover the future abilities of humanoid robots by presenting a variety of integrated research in various scientific and engineering fields, such as locomotion, perception, adaptive behavior, human-robot interaction, neuroscience and machine learning. The book is designed to be accessible and practical, with an emphasis on useful information to those working in the fields of robotics, cognitive science, artificial intelligence, computational methods and other fields of science directly or indirectly related to the development and usage of future humanoid robots. The editor of the book has extensive R&D experience, patents, and publications in the area of humanoid robotics, and his experience is reflected in editing the content of the book.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821