emeraldinsight

Industrial Robot: An International Journal

Gesture-based human-robot interaction using a knowledge-based software platform
Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Ueno,

Article information:

To cite this document:
Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Ueno, (2006) "Gesture-based human-robot interaction using a knowledge-based software platform", Industrial Robot: An International Journal, Vol. 33 Issue: 1, pp.37-49, https://doi.org/10.1108/01439910610638216
Permanent link to this document:
https://doi.org/10.1108/01439910610638216

Downloaded on: 22 November 2018, At: 12:29 (PT)
References: this document contains references to 28 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 556 times since 2006*

Users who downloaded this article also downloaded:

(2007),"Head gesture recognition for hands-free control of an intelligent wheelchair", Industrial Robot: An International Journal, Vol. 34 Iss 1 pp. 60-68 https://doi.org/10.1108/01439910710718469

(2010),"High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition", Industrial Robot: An International Journal, Vol. 37 Iss 2 pp. 137-147 https://doi.org/10.1108/01439911011018911

Access to this document was granted through an Emerald subscription provided by emerald-srm:423484 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Research article

Gesture-based human-robot interaction using a knowledge-based software platform

Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth and H. Ueno
Intelligent Systems Research Division, National Institute of Informatics, Tokyo, Japan

Abstract
Purpose – Achieving natural interactions by means of vision and speech between humans and robots is one of the major goals that many researchers are working on. This paper aims to describe a gesture-based human-robot interaction (HRI) system using a knowledge-based software platform.
Design/methodology/approach – A frame-based knowledge model is defined for the gesture interpretation and HRI. In this knowledge model, necessary frames are defined for the known users, robots, poses, gestures and robot behaviors. First, the system identifies the user using the eigenface method. Then, face and hand poses are segmented from the camera frame buffer using the person's specific skin color information and classified by the subspace method.
Findings – The system is capable of recognizing static gestures comprised of the face and hand poses, and dynamic gestures of face in motion. The system combines computer vision and knowledge-based approaches in order to improve the adaptability to different people.
Originality/value – Provides information on an experimental HRI system that has been implemented in the frame-based software platform for agent and knowledge management using the AIBO entertainment robot, and this has been demonstrated to be useful and efficient within a limited situation.

Keywords Robotics, Man machine interface

Paper type Research paper

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Introduction

Recently, human-robot symbiotic systems have been studied extensively due to the increase of demand of welfare service for the aged and handicapped under the situation of decreasing of the young generation. It is crucial to establish human-robot natural interaction to realize a symbiotic relationship between human and robot. Ueno (2002) proposed a concept of symbiotic information system (SIS) as well as symbiotic robotics system as one of applications where human and robot can communicate with each other in human way using speech and gesture. The objective of SIS is to allow non-expert users, who might not even be able to operate a computer keyboard, to operate robots. Therefore, these robots are necessary to be equipped with natural interfaces using speech and gesture.

Although it has no doubt that the fusion of gesture and speech allows more natural human-robot interaction (HRI), for single modality, gesture recognition can be considered more reliable than speech recognition system as human voices are varies by person to person and large number of data needs to be taken care of by the system. Human speech contains three types of information: who the speaker is, what the speaker said, and how the speaker said it (Fong *et al.*, 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 miss-recognition due to the environmental noise. On the other hand, gestures are expressive, meaningful body motions such as physical movements of the 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 it also varies little for person to person.

Two approaches are commonly used to recognize gestures. One is gloved-based approach that requires wearing of cumbersome contact devices and generally carrying a load of cables that connect the device to a computer (Sturman and Zetler, 1994). Another approach is vision-based technique that does not require wearing any of contact devices with human body part, but uses a set of video cameras and computer vision techniques to interpret gestures (Pavlovic *et al.*, 1997).

This paper concentrates on visual gestures recognition and interpretation in SIS. Though most of the human gestures are made by hands, same hand gestures may have different meaning in different culture. For example, "thumb up" sign means good in US but in the eastern countries like Bangladesh, India, or Pakistan, this gesture considered as

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0143-991X.htm

Industrial Robot
33/1 (2006) 37–49
© Emerald Group Publishing Limited [ISSN 0143-991X]
[DOI 10.1108/01439910610638216]

The authors would like to thank Professor Y. Shirai, Department of Computer Controlled and Mechanical Systems, Osaka University, Japan, for his valuable suggestions and stimulating discussions. The authors would also like to thank to Dr H. Gotoda, National Institute of Informatics, Japan, for his encouragements and discussions.

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

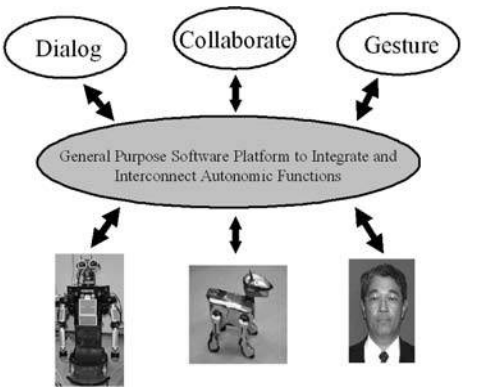
rude sign (Axtell, 1990). Thus, the interpretation of recognized gesture is user-dependent. The recognition process itself is also user-dependent as the skin colors of the hand region and hand shapes are also different by each people. As a result, person identification is the prime factor in order to realize reliable gesture recognition and interpretation.

There are significant amount of researches on hand, arm and facial gesture recognition to control robot or intelligent machine in recent years. Waldherr *et al.* (2000) proposed gesture-based interface for human and service robot interaction. They combined template-based approach and neural network-based approach for tracking a person and recognizing gestures involving arm motion. Watanabe and Yachida (1998) used eigenspaces from multi-input image sequences for recognizing gesture. Single eigenspaces are used for different poses and only two directions are considered in their method. Hu (2003) proposed hand gesture recognition for human-machine interface of robot teleoperation using edge features matching. Rigoll *et al.* (1997) used HMM-based approach for real-time gesture recognition. In their work, features are extracted from the differences between two consecutive images and target images, which are always assumed to be in the center of the input images. Practically it is difficult to maintain such condition, however, Utsumi *et al.* (2002) detected predefined hand pose using hand shape model and tracked hand or face using extracted color and motion. Multiple cameras are used for data acquisition to reduce occlusion problem in their system. But in this process there incurs complexity in computations. Bhuiyan *et al.* (2003, 2004) detected and tracked face and eye for HRI. 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.

In this paper, our face and gesture recognition system for person-centric HRI is based on a knowledge-based software platform called SPAK (software platform for agent and knowledge management), which was developed at our laboratory for intelligent service robots under the internet-based distributed environment (Ampornaramveth and Ueno, 2001). SPAK has been developed to be a platform on which various software components for different robotic tasks can be integrated over a networked environment. It co-ordinates the operation of these components by means of a frame-based knowledge modeling (Minsky, 1974) as shown in Figure 1. SPAK works as a knowledge and data management system, communication channel, intelligent recognizer, intelligent scheduler, and so on. Zhang *et al.* (2004b) has developed a industrial robot arm control systems using SPAK. In that system SPAK works as a communication channel and intelligent robot actions scheduler. Kiatisevi *et al.* (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. It should be noted that since internet communication functions are embedding within the SPAK, developers are not requested to consider complicated communication in developing remote robotic service systems, such as remote robotic control and remote welfare robotic service systems.

This research combines vision and knowledge-based approaches for person-centric HRI so that user can define

Figure 1 Diagram of symbiotic robot system using SPAK knowledge platform



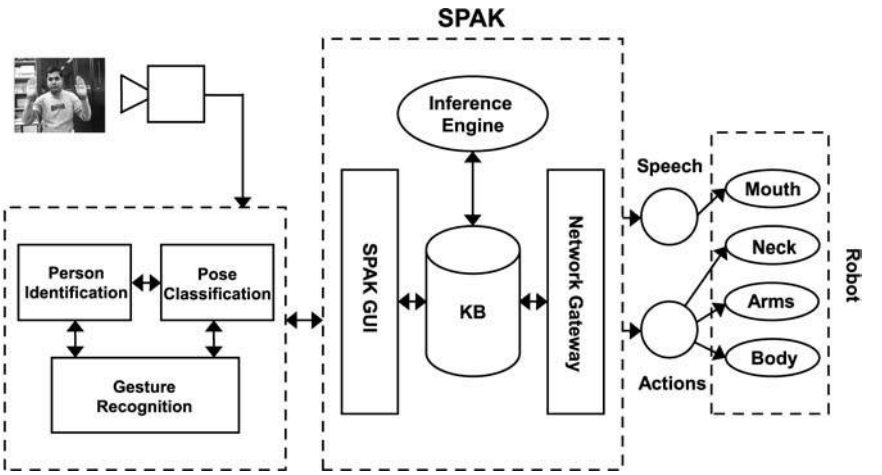
or edit robot behavior according to his/her desire. User can also define or edit the rules for gesture recognition/interpretation in the knowledge database, which also contains information regarding user (profile) to improve reliability and robustness of image classification. For example, the segmented skin regions are more noise free for known person because the probable hand and face poses are segmented using the person-centric threshold values for the YIQ components. Training images are prepared under different illuminations to adapt the system with illumination variation. To achieve better accuracy, this system uses subspace method or separate eigenspaces for hand and face poses classification instead of normal PCA (principal component analysis) method. Dynamic gestures are included in this system by tracing the transition of face and hand poses with the classification of static poses. As an application of this system, we have implemented real-time HRI system using an entertainment robot AIBO.

System architecture

Figure 2 shows the overall architecture of our gesture-based HRI system. The system first detects human face using multiple features and recognizes the face using eigenface method (Turk and Pentland, 1991). Then, using the knowledge of the identified person, face and hand poses are classified and gestures are recognized. The user profile consists of the threshold values for chrominance and luminance components of the skin colors of each known person. Images of face and hand poses are segmented using these color information and classified using the subspace method-based pattern-matching approach. For unknown person, average color information is used and an off-line profile update can be performed later.

After hand and face poses are classified, the static gestures are recognized using frame-based approach. Known gestures are defined as frames in SPAK knowledge base. When the required combination of the pose components is found the corresponding gesture frame is activated. Dynamic gestures are recognized by considering the transitions of the face poses in a sequence of time steps. After the gesture is recognized, the interaction between human and robot is determined by the knowledge modeled also as frame hierarchy in SPAK. Using the received gesture and user information, SPAK processes the facts and activates the corresponding action

Figure 2 Architecture of knowledge-based HRI system by means of gesture



Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

frames to carry out predefined robot actions which may include body movement and speech.

Frame-based knowledge

Knowledge is the theoretical or practical understanding of a subject or a domain. The "frame-based approach" is a knowledge-based problem solving approach based on the so-called, "frame theory", first proposed by 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.

It is well-known that the frame representation systems are currently the primary technology used for large scales knowledge representation in AI (Koller and Pfe.er, 1998). Framed-based approach has been used successfully in many robotic applications (Ueno, 2002). We believe it is quite flexible for realizing human friendly intelligent service robot as well. Our system employs SPAK (Ampornaramveth and Ueno, 2003), 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. This kind of integration is applicable to many robotic applications including industrial robots.

Knowledge model

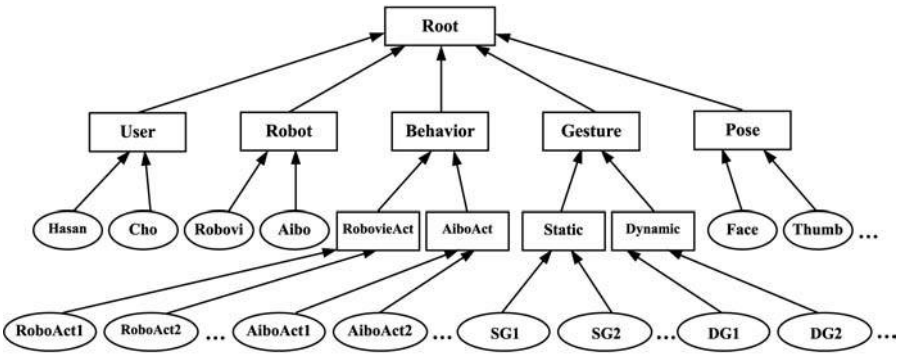
Figure 3 shows the frame hierarchy of the knowledge model for the gesture-based HRI system, organized by the IS_A relationship indicated by arrows connecting upper- and lower-frame boxes. The IS_A relationship is a basic scheme to represent the knowledge in an abstract-specific hierarchy. The values of upper-frame attributes are inherited by lower-frames. An upper-frame in our system describes a group of objects with common attributes. There are upper-frames for user, robot, behavior (robot actions), gesture and pose. An instance of a frame represents a particular object. The user-frame includes instances of all known users (instance "Hasan", "Cho", …), robot frame includes instances of all the robots ("Aibo", "Robovie", …) that are used by 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 gesture, 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 or a face has been recognized. Behavior frames are activated when the predefined conditions are met.

Table I shows the instance-frame "Hasan" under the frame "User". The user-frame attributes include the threshold values for the chrominance and the luminance components of the skin colors of that particular user. In this frame, six slots are defined for the min and max values of the YIQ (Y-luminance and I, Q-chrominance) components. These are used for segmenting skin-like regions from YIQ color space for that person.

This frame-based model can support multiple robots. According to user selection, only the corresponding robot will be activated. Table II shows a sample instance-frame "Aibo" under the class-frame "Robot". In this model, each recognizable pose is treated as an instance-frame under the class frame "Pose". If a predefined pose is classified, then corresponding pose-frame will be activated. Table III shows the frame "Face" of the class-frame "Pose".

The static gestures are defined using the combination of face and hand poses. Table IV shows the example components of an instance-frame "One" of the class-frame "Static gesture". This frame has three slots for the gesture components. If "FACE" and pose "ONE" (raise index finger) are presented in the image and others predefined pose are absent in the image then the value of the slot1 is "FACE", slot2 is "ONE" and slot3 is "Nill". In this combination, gesture "One" is recognized and corresponding frame will be

Figure 3 Frame hierarchy for gesture-based HRI system



Note: SG = static gesture, DG = dynamic gesture

Table I Instance-frame "Hasan" of class-frame "User"

Frame	Hasan
Type	Instance
A-kind-of	User
Has-part	NULL
Semantic-link-from	NULL
Semantic-link-to	NUUL
Slot1	
Name	Y-high
Type	Integer
Slot2	
Name	Y-low
Type	Integer
Slot3	
Name	I-high
Type	Integer
Slot4	
Name	I-low
Type	Integer
Slot5	
Name	Q-high
Type	Integer
Slot6	
Name	Q-low
Type	Integer

Table II Instance frame "Aibo" of class-frame "Robot"

Frame	Aibo
Type	Instance
A-kind-of	Robot
Has-part	NULL
Semantic-link-from	NULL
Semantic-link-to	Behavior

activated. Each robot behavior uses a command or series of commands for a particular task.

Table V shows an example of a behavior-frame for the AIBO robot. It contains slots that are defining the robot

Table III Instance-frame "Face" of class-frame "Pose"

Frame	Face
Type	Instance
A-kind-of	Pose
Has-part	NULL
Semantic-link-from	NULL
Semantic-link-to	Gesture

Table IV Instance-frame "One" of a class-frame "Gesture"

Frame	One
Type	Instance
A-kind-of	Gesture
Has-part	NULL
Semantic-link-from	NULL
Semantic-link-to	Behavior
Slot1	
Name	mFace
Type	Instance
Condition	Any
Argument	FACE
Slot2	
Name	mOne
Type	Instance
Condition	Any
Argument	ONE
Slot3	
Name	mOthers
Type	String
Condition	Equal
Argument	Nil

name, the user name, the gesture name and the function name that activate the robot. This frame is designed for the action "AIBOActStand", and it will be activated if the user is "Hasan" and the gesture is "One" and robot name is "AIBO". The slot4 "OnInstantiate" contains the function name "aibo(PLAY…)" to activate the corresponding actions of the AIBO robot.

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Table V Instance-frame "AiboActStand" of the class-frame "Aibo Behavior"

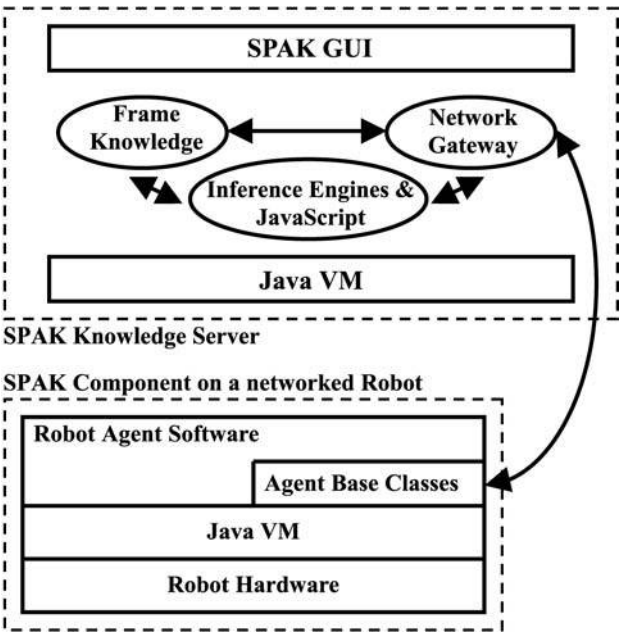
Frame	AIBOActStand
Type	Instance
A-kind-of	AiboAct (behavior)
Has-part	NULL
Semantic-link-from	NULL
Semantic-link-to	NULL
Slot1	
Name	mRobot
Type	Instance
Condition	Any
Argument	Aibo
Slot2	
Name	mUser
Type	Instance
Condition	Any
Argument	Hasan
Slot3	
Name	mGesture
Type	Instance
Condition	Any
Argument	One
Slot4	
Name	OnInstantiate
Type	String
Condition	Any
Argument	Function name

Knowledge management system

The knowledge in this system is maintained by SPAK (SPAK management). SPAK consists of a frame-based knowledge management system and a set of extensible autonomous software agents representing objects inside the environment and supporting HRI and collaborative operation with distributed working environment (Ampornaramveth and Ueno, 2003). It is a Java-based software platform to support knowledge processing and coordination of tasks among several software modules and agents representing the robotic hardware connected on a network.

SPAK consists of the following major components: GUI (graphical user interface), knowledge manager, inference engine, JavaScript interpreter, base class for software agent and network gateway as shown in Figure 4. GUI unit provides the users direct access to the frame-based knowledge. SPAK knowledge manager maintains the frame systems as Java class hierarchy, and performs knowledge conversation to/from XML format. Inference engines verify and process information from external modules, which may result in instantiation or destruction of frame instances in the knowledge manager, and execution of predefined actions. JavaScript interpreter interprets JavaScript code, which is used for defining condition and procedural slots in a frame. It also provides access to rich set of standard Java class libraries that can be used for customizing SPAK to a specific application. Base class for software agent provides basic functionality for developing software agents that reside on network hardware such as robot, a camera, or any software components. Network gateway allows network software agents to access knowledge stored in SPAK.

Figure 4 Relationship between SPAK components on a knowledge server and a network robot



SPAK allows TCP/IP-based communication with other software components in the network and provides knowledge access and manipulation via TCP socket. Figure 4 shows the relationship between SPAK components on a knowledge server and a network robot. In Figure 4, the robot agent communicates with SPAK knowledge server using methods provided by the "Agent Base Class". It may control the robot hardware directly or by communicating with a local native robot control program.

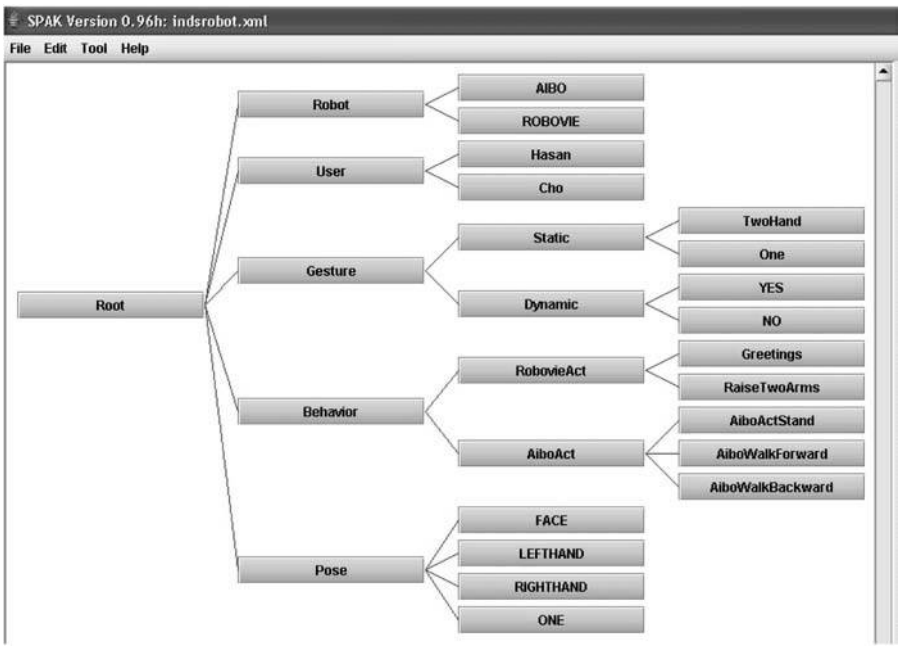
Frame-based knowledge is entered into the SPAK system with full slot information: attributes, conditions and actions. Based on information from remote software components (e.g. pose classification, face recognition, etc.), SPAK inference engine processes facts, instantiate frame instances and carries out the users predefined actions. Figure 5 shows an example SPAK knowledge editor (a part of this system) with the class-frames: robot, user, gesture, behavior (robot actions) and pose. Table VI shows an example of the frame defined in XML format in SPAK for the gesture "One". XML format clearly shows frame structure as well as its contents written by slots can be defined easily (Zhang *et al.*, 2004a, b). A frame is defined between ⟨FRAME⟩ and ⟨/FRAME⟩. NAME refers the frame name. ISA refers to the item of "A-kind-of" relation. ⟨ISINSTANCE⟩ indicates if this is an instance. Multiple slots can be defined between ⟨SLOTLIST⟩ and ⟨/SLOTLIST⟩. Each slot consists of several components, including NAME, TYPE, CONDITION, ARGUMENT, VALUE, REQUIRED, SHARED, etc. The frame system can be created in GUI and interpreted by the SPAK inference engine. The new users, or new poses, or new robot actions can be added by creating corresponding frames in the knowledge base using the SPAK knowledge editor.

Gesture recognition

Images containing faces and hand poses are essential for vision-based HRI. But still it is very difficult to segment face

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Figure 5 Example of SPAK knowledge editor



Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

and hand pose in real time from the color images with clutter background. Human skin color has been used and proven to be an effective feature in many application areas, from face detection to hand tracking. But different people have different skin colors, i.e. chrominance and luminance components are different for different persons in the same lighting conditions. Therefore, person-specific threshold values for the chrominance and luminance components are important for skin-like regions segmentation. This system identifies the person first and then segments hand and face region using person-specific skin color segmentation. In this system, frames for the known users, poses, and gestures are defined in the knowledge base. The user-frame keeps the threshold values for the YIQ (Hasanuzzaman *et al.*, 2004a) components for skin-like region segmentation as shown in Table I. These values are defined from statistical analysis of the skin regions while new user is registered. Since face and two hands may present in the images at the same time, three largest skins like regions are segmented from the input images using skin color information. The subspace method is then used for classifying face and hand poses from these regions. The gestures are interpreted/recognized by SPAK inference engine.

Person identification

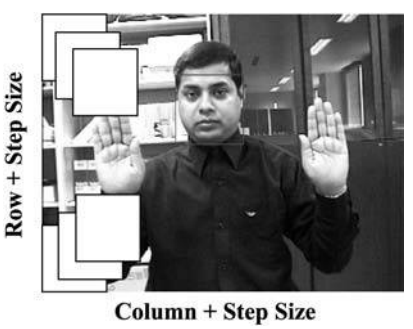
A number of techniques have been developed to detect and recognize faces (Yang *et al.*, 2002). In order to develop a real-time application, robust and efficient face detection algorithm is required. However, face detection from a single image is a challenging task because of variability in scale, location, orientation and pose. Facial expression, occlusion and lighting conditions also change the overall appearance of faces. There are several approaches of face detection, such as knowledge-based (Yang and Huang, 1994), facial features invariant (Sirohey, 1993), template matching (Hasanuzzaman *et al.*, 2004b) and appearance-based (Rowley *et al.*, 1998; Hasanuzzaman *et al.*, 2004c) methods. Our method combines template matching and feature invariant approaches for face detection in order to correctly detect hand poses with face-like elliptical shape. It also uses face template pyramid with different resolutions and orientations. The face templates are scanned on every position of the input image and the matching probability is calculated using minimal Manhattan distance (Hasanuzzaman *et al.*, 2004b). If the minimal Manhattan distance is less than the predefined threshold value, then the searching is accomplished. Two eyes on the upper part of the probable face are located to make sure of the presence of the face (Bhuiyan *et al.*, 2004). If two eyes are found in the probable face areas, then the face area is bounded by a square box with the size of the matched template image. Figure 6 shows the face detection method with example output. For each frame template image sliding starts from the (0, 0) position of the images and progresses it by a given step size from left to right and top to bottom. This process is done until template reaches the end of the input image. This system uses the template images of 50×50, 60×60, 70×70, 80×80, 90×90, 100×100 and 110×110 dimensions for face detection.

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. The detected face is scaled to be a square image with 60×60 dimension and converted to be a gray image. This face pattern is classified using the eigenface method (Turk and Pentland, 1991), whether it belongs to known person or unknown person. The face recognition method uses five face classes: normal face or frontal face (P1), right directed face (P2), left directed face (P3), up state face (P4) and down state face (P5) in training images as shown in Figure 7 (top row). The eigenvectors are calculated from the known persons face images for each face class and k-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

Table VI Example of XML code for defining a gesture-frame "One"

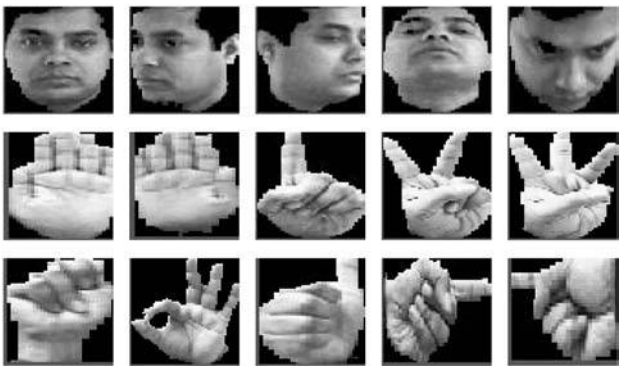
```
⟨FRAME⟩
   ⟨NAME⟩One⟨/NAME⟩
   ⟨ISA⟩Gesture⟨/ISA⟩
   ⟨ISINSTANCE⟩FALSE⟨/ISINSTANCE⟩
⟨SLOTLIST⟩
⟨SLOT⟩
   ⟨NAME⟩mFace⟨/NAME⟩
   ⟨TYPE⟩TYPE_INSTANCE⟨/TYPE⟩
   ⟨CONDITION⟩COND_ANY⟨/CONDITION⟩
   ⟨ARGUMENT⟩FACE⟨/ARGUMENT⟩
   ⟨VALUE⟩⟨/VALUE⟩
   ⟨REQUIRED⟩TRUE⟨/REQUIRED⟩
   ⟨SHARED⟩TRUE⟨/SHARED⟩
   ⟨UNIQUE⟩TRUE⟨/UNIQUE⟩
⟨/SLOT⟩
⟨SLOT⟩
   ⟨NAME⟩mOne⟨/NAME⟩
   ⟨TYPE⟩TYPE_INSTANCE⟨/TYPE⟩
   ⟨CONDITION⟩COND_ANY⟨/CONDITION⟩
   ⟨ARGUMENT⟩ONE⟨/ARGUMENT⟩
   ⟨VALUE⟩⟨/VALUE⟩
   ⟨REQUIRED⟩TRUE⟨/REQUIRED⟩
   ⟨SHARED⟩TRUE⟨/SHARED⟩
   ⟨UNIQUE⟩TRUE⟨/UNIQUE⟩
⟨/SLOT⟩
⟨SLOT⟩
   ⟨NAME⟩mOthers⟨/NAME⟩
   ⟨TYPE⟩TYPE_STR⟨/TYPE⟩
   ⟨CONDITION⟩COND_EQ⟨/CONDITION⟩
   ⟨ARGUMENT⟩Nil⟨/ARGUMENT⟩
   ⟨VALUE⟩⟨/VALUE⟩
   ⟨REQUIRED⟩TRUE⟨/REQUIRED⟩
   ⟨SHARED⟩TRUE⟨/SHARED⟩
   ⟨UNIQUE⟩TRUE⟨/UNIQUE⟩
⟨/SLOT⟩
⟨/SLOTLIST⟩
⟨/FRAME⟩
```

Figure 6 Example of the face detection method



eigenspaces. If the minimal Euclidian is less than the predefined threshold value then person is known, otherwise unknown. The detail of this method is described in our previous research (Hasanuzzaman *et al.*, 2004c).

Figure 7 Examples of training images



Face and hand poses segmentation

Face and hand poses are segmented using person-specific skin color information. In this paper, YIQ (Y is luminance of the color and I, Q are chrominance of the color) color representation system is used for skin-like region segmentation because it is typically used in video coding and provides an effective use of chrominance information for modeling the human skin color (Bhuiyan *et al.*, 2003; Dai and Nakano, 1996). The RGB images taken by the video camera are converted to YIQ color representation system. Skin color region is determined by applying threshold values ((Y_Low $<$ Y $<$ Y_High) && (I_Low $<$ I $<$ I_High) && Q_Low $<$ Q $<$ Q_High)) (Hasanuzzaman *et al.*, 2004a). We have selected fixed range of thresholds values after statistical analysis of human skin regions rather than dynamical updateable thresholds values. This color-based segmentation is simple, fast and works well in specific rooms. This method uses different threshold values for the YIQ components for different persons. The ranges of threshold values are selected from the knowledge base for known person as shown in Table I.

The thresholds are predetermined from the histogram of YIQ components of the skin region and included Y_High, Y_Low, I_High, I_Low, Q_High and Q_Low as the ranges of threshold values of the Y, I, Q components in the knowledge base with corresponding person identity. After identifying the person this values are used for skin region segmentation. Probable hands and face regions are isolated from the image with the three largest connected regions of skin-colored pixels. The notation of pixel connectivity describes a relation between two or more pixels. In order to consider two pixels to be connected, their pixel values must both be from the same set of values V (for binary images v is 1). General connectivity can either be based on 4- or 8-connectivity. In the case 4-connectivity, it does not compare the diagonal pixels but 8-connectivity compares the diagonal positions pixels considering 3 × 3 matrix and more noise free than 4-connectivity. In this system, 8-pixels neighborhood connectivity is employed that was developed by our group member (Bhuiyan *et al.*, 2004).

In order to remove the false regions from the segmented blocks, smaller connected regions are assigned by the values of black-color (R = G = B = 0). As a result, after thresholding the segmented image may contain some holes in the three largest skin-like regions. In order to remove noises and holes, segmented images are filtered by morphological dilation and erosion operations. The dilation operation is used

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

to fill the holes and the erosion operations are applied to the dilation results to restore the shape. If the person shirt's color is similar to skin color then segmentation accuracy is very poor. If the person wears short sleeves T-shirt then it needs to separate hand palm from arm. This system assumes the person wearing full shirt with non-skin color. Normalization is done to scale the image to match with the size of the training image and convert the scaled image to gray image (Hasanuzzaman *et al.*, 2004a).

Pose classification using subspace method

The main idea of the subspace method is similar to the PCA that is to find the vectors that best account for the distribution of target images within the entire image space. In the normal PCA method, eigenvectors are calculated from training images that include all the poses or classes. In the subspace method, training images are grouped for face and hand poses separately. In subspace method target image is projected on each subspace separately.

The procedure of face and hand pose classification using subspace method includes following operations. The meanings of symbols are shown in Table VII.

- Prepare noise free version of predefined face and hand poses corresponding training images $T_j^{(i)}(N \times N)$, where j is number training images of ith class and $j = 1, 2 \ldots, M$. Figure 7 shows the example training image classes: frontal face, right directed face, left directed face, up directed face, down directed face, left hand palm, right hand palm, raised index finger, raised index and middle finger to form "V" sign, raised index, middle and ring fingers, fist up, make circle using thumb and fore fingers, thumb up, point left by index finger and point right by index finger are defined as pose P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14 and P15, respectively.
- For each class, calculate eigenvectors $(u_m^{(i)})$ using Turk and Pentland (1991) technique and chose k-number of eigenvectors $(u_k^{(i)})$ corresponding to the highest eigenvalues to form principal components for that class. These vectors for each class define the subspace of that group.
- Calculate corresponding distribution in k-dimensional weight space for the known training images by projecting them onto the subspaces (eigenspaces) of the corresponding class and determined the weight vectors $(\Omega_l^{(i)})$, using equations (1) and (2).

Table VII List of symbols

Symbols	Meanings
$T_j^{(i)}$	Training images for ith class
$u_m^{(i)}$	mth Eigenvectors for ith class
$\Omega_l^{(i)}$	Weight vector for ith class
$\omega_k^{(i)}$	Element of weight vector for ith class
Φ_l	Average image for ith class
$s_l^{(i)}$	lth Known image for ith class
ε	Euclidean distance among weight vectors
$\varepsilon_l^{(i)}$	Element of Euclidean distance among weight vectors for ith class

$$\omega_k^{(i)} = \left(u_k^{(i)}\right)^{\mathrm{T}} \left(s_l^{(i)} - \Phi_i\right) \quad (1)$$

$$\Omega_l^{(i)} = \left[\omega_1^{(i)}, \omega_2^{(i)}, \ldots, \omega_k^{(i)}\right] \quad (2)$$

where, average image of ith class

$$\Phi_i = \frac{1}{M} \sum_{n=1}^{M} T_n$$

and $s_l^{(i)}(N \times N)$ is lth known images of ith class.

- Each segmented skin-region is treated as individual input image and transformed into eigenimage components and calculate a set of weight vectors $(\Omega^{(i)})$ by projecting the input image onto each of the subspace as equations (1) and (2).
- Determine if the image is a face pose or other predefined hand pose based on minimum Euclidean distance among weight vectors using equations (3) and (4):

$$\varepsilon_l^{(i)} = \|\Omega^{(i)} - \Omega_l^{(i)}\| \quad (3)$$

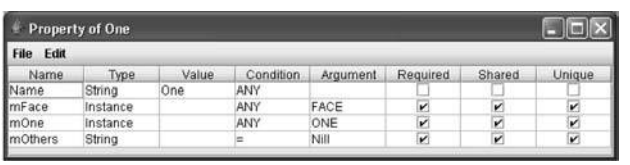
$$\varepsilon = \left[\varepsilon_1^{(1)}, \varepsilon_2^{(1)}, \ldots, \varepsilon_l^{(i)}\right] \quad (4)$$

If $\min\{\varepsilon\}$ is lower than predefined threshold then its corresponding pose is identified. For exact matching ε should be zero but for practical purposes this method uses a threshold value obtained from experiment. If the pose is identified then corresponding frame will be activated.

Recognizing gesture

Gesture recognition is the process by which gestures made by the user are identified in the system. There are static gesture and dynamic gesture. The recognition of gesture is carried out in two phases. In the first phase, face and hand poses are classified from the each captured image frame using the method described in previous section. Then sequence of poses and combination of poses are analyzed to identify the occurrence of gesture. Interpretation of identified gesture is user-dependent since the meaning of the gesture may differ from person to person based on their culture. Our knowledge management system mapped user-gesture-robot actions. For example, if the system recognized the user as "Hasan" and gesture as "One" then "AIBO" robot action is "STAND UP". But for another case, if user is "Cho" and gesture is "One", then the action of "AIBO" robot is "WALK FORWARD". To accommodate different user's desires, our person-centric gesture interpretation is implemented using frame-based approach. Table IV shows the example components of gesture "One". The gesture frame is defined using three slots corresponding to pose recognition results of three skin-regions. Figure 8 shows the gesture frame "One" in SPAK. If pose "ONE" (raised index finger) and "FACE" present in the input image and other predefined pose is absent (Nil) then recognizes it as gesture "One". Our current system recognizes

Figure 8 Sample frame for the gesture "One"



Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

13 gestures: 11 static gestures and 2 dynamic facial gestures as shown in Table VIII. It is possible to recognize more gestures including new poses and new rules using this system. New pose can be included in the training image database and corresponding frame can be defined in the knowledge base to interpret the gesture.

Static gesture recognition

Static gestures are recognized using frame-based system with the combination of the pose classification results of the three skin-like regions at a particular time. For examples, if left hand palm, right hand palm and one face present in the input image then recognizes it as "TwoHand" gesture and corresponding frame will activate. The user predefines these frames in knowledge base as shown in Table VI. The system maintains frames with necessary attributes (gesture components, gesture name) for all predefined gestures. Gesture components are the face and hand poses. If the pose is identified then pose name is feed to SPAK and corresponding instance frame of the pose-frame will be activated. Figure 8 shows an example gesture frame ("One") with all the components that is defined using SPAK. If pose "ONE" (raised index finger) and face present in the images then recognizes it as gesture "One". Similarly other static gestures are recognized.

Dynamic gesture recognition

Dynamic gesture is the gesture that uses motion of hand or body to emphasize or help to express a thought or feelings such as "NO" (shake face left-right), "YES" (shakes face up-down), "Come" (beaconing by hand), "Bye-bye" (waving by hand), etc. It is difficult to visually recognize dynamic gesture due to large variations in the speed of position change of the physical objects that describe the gesture. In our system dynamic gestures are recognized from the specific motion pattern or pose sequences in time steps and can be defined using state transition diagram. This system recognizes two dynamic facial gestures considering the transition of the face poses in a sequence of time steps. If human face shakes left to right or right to left, then it is recognized as "NO" gesture. If human face shakes up and down or down and up, then it is recognized as "YES" gesture. This method uses a three-layers queue (FIFO) that holds the last three results of the detected face poses. This method defines five specific face poses: frontal face (NF), right-rotated face (RF), left-rotated face (LF), up position face (UF) and down position face (DF) as shown in the top row of Figure 7 from left to right. For every image frame, face pose is classified using subspace method. If pose is classified as predefined face pose then it is added to the three-layer queue. If the classified pose value is same as previous frames then queue values will remain unchanged. From the combination of three-layers queue values this method determine the gesture. For example, if the queues values are {UF, NF, DF} or {DF, NF, UF} pose sets then it is recognized as "YES" gesture. Similarly, if the queues values are {RF, NF, LF} or {LF, NF, RF} pose sets then it is recognized as "NO" gesture. After a specific time period the queue values are refreshed. Figure 9 shows example face sequences for dynamic gestures "YES" (Figure 9(a)) and "NO" (Figure 9(b)). Figure 10 shows the state transition diagram. If gesture is recognized then corresponding gesture frame will be activated.

Table VIII Three segments combination and corresponding gesture(X = absence of predefined hand poses or face poses)

Gesture components			Gesture names
Face	Left hand palm	Right hand palm	TwoHand
Face	Right hand palm	X	RightHand
Face	Left hand palm	X	LeftHand
Face	Index finger raise	X	One
Face	Form V sign with index and middle finger	X	Two
Face	Index, middle and ring fingers raise	X	Three
Face	Thumb up	X	ThumbUp
Face	Make circle using thumb and index finger	X	OK
Face	Fist up	X	FistUp
Face	Point left by index finger	X	PointLeft
Face	Point right by index finger	X	PointRight
Shakes face left and right or right and left			NO
Shakes face up and down or down and up			YES

Experiments and discussions

Experiment setup

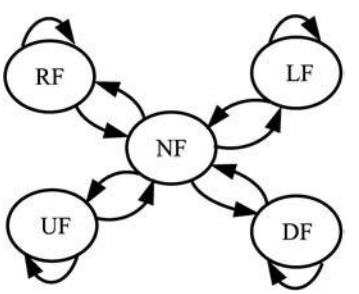
This system uses a standard video camera for data acquisition. Each captured image is digitized into a matrix of 320 × 240 pixels with 24-bit color. The recognition approach has been tested with real world HRI system using an entertainment

Figure 9 Example of dynamic gesture sequences

(a) Gesture "YES" {UF, NF, DF}

(b) Gesture "NO" {RF, NF, LF}

Figure 10 State transition diagram of dynamic gesture sequences



Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

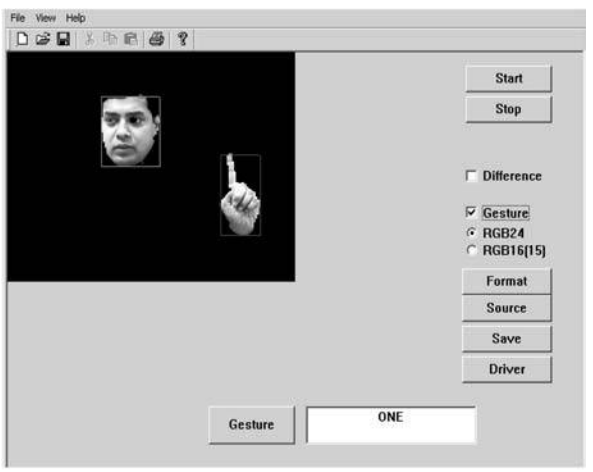
robot AIBO (developed by Sony). First, the system is trained using the training images for 15 poses (five face poses and ten hand poses) of seven persons. All the training images are 60 × 60 pixels gray images. The training images consist of 2,100 images for 15 poses of 7 person; 140 images for each pose of 7 persons. In the training phase, this system generates eigenspaces and feature vectors for the known users and known hand poses. The threshold values for the YIQ components are also defined for each known user in the knowledge base during the training phase. This system is tested for real-time input images as well as static images. This system is also tested for the ASL characters classification.

Gesture recognition results

The sample visual output of gesture recognition system is shown in Figure 11(a). It shows the gesture name at the bottom text box corresponding to matched gesture ("ONE"). In the case of no match, it shows "no matching found" in this text box.

In this study, we have compared the performance of the general PCA and the pose-specific-subspace (for each pose one PCA) method for pose classification. For this experiments we have trained the system using 2,100 training images for 15 poses of 7 persons (140 images for each pose of seven persons). A total of seven individuals were asked to act for the predefined face and hand poses in front of the camera and the sequence of images were saved as individual image frame. Then each image frame is tested using the normal PCA and the subspace method. The threshold value (for minimal Euclidian distance) for the pose classifier is selected so that all the poses are classified. Tables IX and X show the confusion matrix using the PCA method and the subspace method for 15 poses (as shown in Figure 7) of 7 persons. In this confusion matrix the diagonal elements represent the correct recognition of each pose. The column represents the classification results and the row represents the input image class.

Figure 11 Sample output of gesture based human-robot (AIBO) interaction



(a) Sample visual output ("One")

(b) AIBO STAND-UP

Table IX Confusion matrix using PCA method

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1	122	2	1	3	2	0	0	0	0	0	0	0	0	0	0
P2	0	142	0	1	3	0	0	0	0	1	0	0	0	8	0
P3	0	0	139	0	0	0	0	0	0	0	0	1	0	0	0
P4	2	0	0	144	1	0	0	0	0	0	0	0	0	0	0
P5	7	0	0	0	120	0	1	0	0	0	0	0	0	0	0
P6	0	0	0	0	9	128	3	0	0	0	0	0	0	0	0
P7	1	1	1	0	0	0	137	0	0	0	0	0	0	0	0
P8	0	0	0	0	0	0	0	138	0	0	0	2	0	0	0
P9	0	0	0	0	0	0	0	2	126	9	0	1	0	2	0
P10	0	0	0	0	1	0	0	1	2	133	0	1	0	2	0
P11	6	0	8	4	0	0	0	0	0	0	131	1	0	0	0
P12	0	1	0	2	0	0	1	0	0	0	0	146	0	0	0
P13	0	1	1	0	0	0	0	6	2	0	0	0	140	0	0
P14	1	2	0	5	0	0	0	0	0	0	0	0	0	122	0
P15	0	0	0	1	0	0	0	0	0	0	0	0	0	0	149
Total	139	149	150	160	136	128	142	147	130	143	131	152	140	134	149

Table XI shows the comparison of precision and recall rate of the subspace method and the standard PCA method for face and hand poses classification. From the results, we conclude that precision and recall rates are higher in the subspace method and wrong classification rates are lower than the standard PCA method. Wrong classification occurred due to orientation and intensity variation.

Note that, the accuracy of the gesture recognition system depends on the accuracy of the pose classification unit. For example, in some cases, pose 9 ("V sign") is present in the input image but the pose classification method failed to classify it correctly and classified it as pose 8 ("raised index finger") due to the variation of orientation, then the gesture recognition output is "One". The accuracy of the dynamic gesture recognition also depends on the accuracy of the face pose classification. Table XII shows the success rate of dynamic gestures recognition. This test is done on six image sequences and each image sequence is consists of 485 image frames. For correct recognition, at least each face pose should be at the stable state for one image frame time.

This system also can classify 26 ASL characters (A to Z) (ASL American Sign Language Browser http://commtechlab.msu.edu/sites/aslweb/browser.htm). But sign words (vocabulary) recognition, requires to track the hand motion as well as the position with respect to other important parts of the body such as the head, chest and shoulders. Vision-based

Table X Confusion matrix using subspace method

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
P1	127	0	0	3	0	0	0	0	0	0	0	0	0	0	0
P2	3	152	0	0	0	0	0	0	0	0	0	0	0	0	0
P3	0	0	139	0	1	0	0	0	0	0	0	0	0	0	0
P4	1	0	0	146	0	0	0	0	0	0	0	0	0	0	0
P5	1	0	0	1	126	0	0	0	0	0	0	0	0	0	0
P6	0	0	0	0	0	132	2	1	0	0	0	5	0	0	0
P7	0	0	0	0	0	0	140	0	0	0	0	0	0	0	0
P8	0	0	0	0	0	0	0	138	0	0	0	2	0	0	0
P9	0	0	0	0	0	0	0	2	132	3	0	1	0	2	0
P10	0	0	0	0	0	0	0	0	3	133	0	2	1	1	0
P11	0	0	0	0	0	1	0	0	0	0	139	0	8	2	0
P12	0	0	0	0	0	1	1	0	0	0	0	147	0	2	0
P13	0	0	0	0	0	0	1	5	2	0	0	0	142	0	0
P14	0	0	0	0	0	0	0	0	0	0	3	0	0	127	0
P15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150
Total	132	152	139	150	127	133	144	146	137	136	142	157	151	134	150

Table XI Comparison of subspace method and PCA method

Pose number	Precision (percent) Subspace	PCA	Recall (percent) Subspace	PCA
P1	96.21	90.37	97.69	93.84
P2	100	96.59	98.06	91.61
P3	100	93.28	99.28	99.28
P4	97.33	92.30	99.31	97.95
P5	99.21	90.90	98.43	93.75
P6	100	100	94.28	91.42
P7	97.22	96.47	100	97.85
P8	95.17	94.52	98.57	98.57
P9	97.77	97.67	94.28	90
P10	97.81	93.05	95	95
P11	100	100	92.66	87.33
P12	96.71	96.68	98	97.33
P13	99.31	100	94.66	93.33
P14	94.89	93.28	97.69	93.84
P15	100	100	100	99.33

Table XII Evaluation of dynamic gesture recognition

Image sequence	Success gesture Yes	No	Un-success gesture	Average success rate (percent)
Seq1	19	9	2	93.33
Seq2	18	20	0	100
Seq3	23	12	1	97.22
Seq4	16	20	2	94.73
Seq5	35	11	3	93.87
Seq6	18	14	2	94.11

sign language recognition is still difficult and it is the future aim of gesture recognition society. This system was tested with 3,400 ASL character poses of different persons. The success rate for ASL character recognition is about 95 percent, but it is still difficult to distinguish all characters only using pose classification. For example, character I is very similar to J and X is very similar to Z considering static shape but there are a transitions from I to J and X to Z (dynamic gesture). For correct recognition of such ASL characters we need to analyze the motion also.

The proposed face detection method in this paper is robust against background, motion and distance, but this method has a larger computational cost that is the bottleneck for real time HRI. Three factors directly affect on computation costs: step size, template images dimension and the number template images. If step size is 1, for sliding one template on the whole image, the number of comparison is 46,800 (60×60, 320×240) where 60×60 is template image dimension and 320×240 is input image dimension. In similar cases if step sizes are 2, 3, 4, 5 then numbers of comparisons are 11,700, 5,220, 2,925, and 1,872, respectively. If the template image dimension increases then reduces the computation cost but in that case small faces are ignored. The computation cost also increases if the number of template images increases. There are many ways to reduce the processing time for face detection: such as motion areas segmentation and human skin areas segmentation. In this work, we use human skin areas segmentation with reasonable step size. This system considers that the robots will operate in a room with fixed lighting condition that is why person centric threshold values (for YIQ components) are used for skin-regions segmentation.

In our previous research (Hasanuzzaman *et al.*, 2004c), we found that the accuracy of frontal face recognition is better than up, down and more left right directed faces. In this system we prefer frontal and a small left-right rotated face only for person identification. We have tested this face recognition method for 680 faces of seven persons, where two are female. The average precision for face recognition is about 93 percent and recall rate is about 94.08 percent.

Knowledge-based human-robot interaction

Since 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 (or gesture name for dynamic gestures) are sent to the SPAK for decisions making and the robot activation. According to gesture and user identity knowledge module generates executables code for robot actions. The robot then follows speech and body actions commands. For example, a user selects the robot AIBO for interaction. The gesture recognition module recognizes the gesture is "One" and the face recognition module identifies the person as "Hasan". In this combination SPAK activates "AIBOActStand" frame corresponding the user "Hasan" and gesture "One" as shown in Figure 12. AIBO robot then "STAND UP" according to predefined action for this combination as shown in Figure 11(b). But for another user same gesture may be used for another action of AIBO. Suppose user "Cho" defines the AIBO action to be "WALK FORWARD" for gesture "One", i.e. if user is "Cho", gesture is "One" and robot name is AIBO then "AIBOActWalkForward" frame will be activated and AIBO will start to walk forward. In similar way, the user can design AIBO action frames according to his/her desires. This example demonstrates that the same gesture

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Figure 12 Frame for AIBO "STAND UP" action for user "Hasan"

is used for different meaning for different persons. The user may define different actions for the same gesture.

The actions of the AIBO are: "STAND UP", "WALK FORWARD", "WALK BACKWARD", "MOVE RIGHT", "MOVE LEFT" "KICK", "SIT" and "LIE" in accordance with gesture "One", "Two", "Three", "PointLeft", "PointRight", "RightHand" "TwoHand", and "LeftHand", respectively for user "Hasan". In the case of user "Cho" the actions of the AIBO are: "STAND UP", "WALK FORWARD", "WALK BACKWARD", "MOVE RIGHT", "MOVE LEFT", "KICK", "SIT", and "LIE" in accordance with gesture "TwoHand", "One", "Two", "RightHand", "LeftHand", "Three", "FistUp", and "ThumbUp", respectively. Figure 11 shows example output of human-robot (AIBO) interaction based on gesture, where AIBO is "STAND UP" for human gesture "One".

The above scenario demonstrates how the systems accounts for the fact that same gesture is used for different meanings and several gestures are used for the same meanings for different persons. The user can design new actions according to his/her desires and can design corresponding knowledge frames using SPAK to implement their desired actions.

It is possible to implement complex command like complete sentence, such as, "Who is he", "Go and Open the Door", "Give me a glass of Water", etc. For complex command we need to make series of predefined gesture commands orderly with a start and end gesture. For "Go and Open the Door" command the operator should made the gesture "Go" first, then pointing gesture to direct the door and finally use "Open the Door" (Hand held as if grasping the handle of a door) gesture. It is possible to implement the complex command using knowledge-based software platform. The complex gesture-frame will be activated if corresponding gestures are recognized sequentially. The future task of our team is to develop such kind of gesture that is helpful for elderly/handicapped. In the proposed system we are not training elderly to learn gestures. Instead, the elderly/deaf people are supposed to know in advance gestures for accomplishing command.

Conclusions

This paper describes a gesture-based HRI system using a knowledge-based software platform. The frame-based knowledge modeling is quite flexible and powerful for gesture interpretation and person-centric HRI. Human skin-color (luminance and chrominance components) differs from person to person so person-centric threshold values for YIQ components are very useful for skin regions segmentation. This system uses separate eigenspaces for face and hand poses so it is more reliable than the normal PCA-based method. In addition, with gesture recognition this system is also capable to identify persons. By integrating with knowledge-based software platform, gesture-based person-centric HRI system has also been successfully implemented in this paper using AIBO. In this system, the user can define or update the rules or condition for gesture recognition/interpretation and the robot behaviors corresponding to his/her gestures.

Face recognition with gesture recognition will help us to develop person adaptive gesture recognition system for human-robot interface. Person-centric gesture should be applicable for culture adaptable gesture interpretation and operator specific industrial robot control. The future aim is to make the system more robust, dynamically adaptable to new user and new gestures for interaction with different robots such as AIBO, ROBOVIE, SCOUT, etc. Our ultimate goal is to establish a human-robot symbiotic society so that they can share their resources and work cooperatively with human beings.

References

Ampornaramveth, V. and Ueno, H. (2001), "Software platform for symbiotic operations of human and networked robots", *NII Journal*, Vol. 3 No. 1, pp. 73-81.

Ampornaramveth, V. and Ueno, H. (2003), "SPAK: software platform for agents and knowledge systems in symbiotic robots", *IEICE Transactions on Information & Systems*, Vol. E86-D No. 3, pp. 1-10.

Axtell, R.E. (1990), *Gestures: The Do's and Taboos of Hosting International Visitors*, Wiley, New York, NY.

Bhuiyan, M.A., Ampornaramveth, V., Muto, S.Y. and Ueno, H. (2003), "Face detection and facial feature localization for human-machine interface", *NII Journal*, Vol. 5 No. 1, pp. 25-39.

Bhuiyan, M.A., Ampornaramveth, V., Muto, S. and Ueno, H. (2004), "On tracking of eye for human-robot interface", *International Journal of Robotics and Automation*, Vol. 19 No. 1, pp. 42-54.

Dai, Y. and Nakano, Y. (1996), "Face-texture model based on SGLD and its application in face detection in a color scene", *Pattern Recognition*, Vol. 29 No. 6, pp. 1007-17.

Fong, T., Nourbakhsh, I. and Dautenhahn, K. (2003), "A survey of socially interactive robots", *Robotics and Autonomous System*, Vol. 42 Nos 3-4, pp. 143-66.

Hasanuzzaman, Md., Ampornaramveth, V., Zhang, T., Bhuiyan, M.A., Shirai, Y. and Ueno, H. (2004a), "Real-time vision-based gesture recognition for human-robot interaction", *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO), China*, pp. 379-84.

Hasanuzzaman, Md., Zhang, T., Ampornaramveth, V., Bhuiyan, M.A., Shirai, Y. and Ueno, H. (2004b), "Gesture recognition for human-robot interaction through a knowledge based software platform", *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR 2004) Portugal, LNCS 3211,* Springer-Verlag, Berlin, Heidelberg, 1, pp. 5300-537.

Hasanuzzaman, Md., Zhang, T., Ampornaramveth, V., Bhuiyan, M.A., Shirai, Y. and Ueno, H. (2004c), "Face and gesture recognition using subspace method for human-robot interaction", paper presented at Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia LNCS 3331, 1, Tokyo, pp. 369–76.

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

Hu, C. (2003), "Gesture recognition for human-machine interface of robot teleoperation", paper presented at International Conference on Intelligent Robots and Systems, pp. 1560-5.

Kiatisevi, P., Ampornaramveth, V. and Ueno, H. (2004), "A distributed architecture for knowledge-based interactive robots", *Proceedings of the 2nd International Conference on Information Technology for Application (ICITA'2004)*, pp. 256-61.

Koller, D. and Pfe.er, A. (1998), "Probabilistic frame-based systems", *Proceedings of the 15th National Conference on AI (AAAI-98)*, pp. 580-7.

Minsky, M. (1974), "A framework for representing knowledge", Memo 306, MIT-AI Laboratory.

Pavlovic, V.I., Sharma, R. and Huang, T.S. (1997), "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. 677-95.

Rigoll, G., Kosmala, A. and Eickeler, S. (1997), "High performance real-time gesture recognition using hidden Markov models", *Proceedings Gesture and Sign Language in Human Computer Interaction, International Gesture Workshop, Germany*, pp. 69-80.

Rowley, H.A., Baluja, S. and Kanade, T. (1998), "Neural network-based face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 23 No. 1, pp. 23-38.

Sirohey, S.A. (1993), "Human face segmentation and identification", Technical Report CS-TR-3176, pp. 1-33, University of Maryland.

Sturman, D.J. and Zetler, D. (1994), "A survey of glove-based input", *IEEE Computer Graphics and Applications*, Vol. 14, pp. 30-9.

Turk, M. and Pentland, A. (1991), "Eigenface for recognition", *Journal of Cognitive Neuroscience*, Vol. 3 No. 1, pp. 71-86.

Ueno, H. (2002), "A knowledge-based information modeling for autonomous humanoid service robot", *IEICE Transactions on Information & Systems*, Vol. E85-D No. 4, pp. 657-65.

Utsumi, A., Tetsutani, N. and Igi, S. (2002), "Hand detection and tracking using pixel value distribution model for multiple-camera-based gesture interactions", *Proceedings of the IEEE Workshop on Knowledge Media Networking (KMN'02)*, pp. 31-6.

Waldherr, S., Romero, R. and Thrun, S. (2000), "A gesture based interface for human-robot interaction", *Journal of Autonomous Robots*, pp. 151-73.

Watanabe, T. and Yachida, M. (1998), "Real-time gesture recognition using eigenspace from multi-input image sequences", *System and Computers in Japan*, Vol. J81-D-II, pp. 810-21.

Yang, G. and Huang, T.S. (1994), "Human face detection in complex background", *Pattern Recognition*, Vol. 27 No. 1, pp. 53-63.

Yang, M-H., Kriegman, D.J. and Ahuja, N. (2002), "Detection faces in images: a survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 24 No. 1, pp. 34-58.

Zhang, T., Ampornaramveth, V., Kiatisevi, P., Hasanuzzaman, Md. and Ueno, H. (2004a), "Knowledge-based multiple robots coordinative operation using software platform", *Proceedings of the 6th Joint Conference on Knowledge-Based Software Engineering (JCKBSE), Russian*, pp. 149-58.

Zhang, T., Hasanuzzaman, Md., Ampornaramveth, V., Kiatisevi, P. and Ueno, H. (2004), "Human-robot interaction control for industrial robot arm through software platform for agents and knowledge management", *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC 2004), Netherlands*, pp. 2865-70.

Corresponding author

Md. Hasanuzzaman can be contacted at: hzaman@grad.nii.ac.jp

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**

This article has been cited by:

1. Gilbert Tang, Phil Webb. 2018. The Design and Evaluation of an Ergonomic Contactless Gesture Control System for Industrial Robots. *Journal of Robotics* **2018**, 1-10. [Crossref]
2. Derek McColl, Alexander Hong, Naoaki Hatakeyama, Goldie Nejat, Beno Benhabib. 2016. A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI. *Journal of Intelligent & Robotic Systems* **82**:1, 101-133. [Crossref]
3. Gilbert Tang, Seemal Asif, Phil Webb. 2015. The integration of contactless static pose recognition and dynamic hand motion tracking control system for industrial human and robot collaboration. *Industrial Robot: An International Journal* **42**:5, 416-428. [Abstract] [Full Text] [PDF]
4. Jens Lambrecht, Martin Kleinsorge, Martin Rosenstrauch, Jörg Krüger. 2013. Spatial Programming for Industrial Robots through Task Demonstration. *International Journal of Advanced Robotic Systems* **10**:5, 254. [Crossref]
5. Jinyung Jung, Takayuki Kanda, Myung-Suk Kim. 2013. Guidelines for Contextual Motion Design of a Humanoid Robot. *International Journal of Social Robotics* **5**:2, 153-169. [Crossref]
6. Derek McColl, Zhe Zhang, Goldie Nejat. 2011. Human Body Pose Interpretation and Classification for Social Human-Robot Interaction. *International Journal of Social Robotics* **3**:3, 313-332. [Crossref]
7. Md. Hasanuzzaman, Tetsunari Inamura. Adaptation to new user interactively using dynamically calculated principal components for user-specific human-robot interaction 164-169. [Crossref]
8. Salma Begum, Md. Hasanuzzaman. Computer Vision-based Bangladeshi Sign Language Recognition System 414-419. [Crossref]
9. Zhi-Hong Mao, Heung-No Lee, R.J. Sclabassi, Mingui Sun. 2009. Information Capacity of the Thumb and the Index Finger in Communication. *IEEE Transactions on Biomedical Engineering* **56**:5, 1535-1545. [Crossref]
10. J.B. Gomez, F. Prieto, T. Redarce. Towards a mouth gesture based laparoscope camera command 35-40. [Crossref]
11. Goldie Nejat, Maurizio Ficocelli. Can I be of assistance? The intelligence behind an assistive robot 3564-3569. [Crossref]
12. Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Gotoda, Y. Shirai, H. Ueno. 2007. Adaptive visual gesture recognition for human–robot interaction using a knowledge-based software platform. *Robotics and Autonomous Systems* **55**:8, 643-657. [Crossref]

Downloaded by Imperial College London At 12:29 22 November 2018 (PT)