Knowledge-Based Person-Centric Human-Robot Interaction Using Facial and Hand Gestures

Md. Hasanuzzaman*, T. Zhang*, V. Ampornaramveth*, H. Gotoda *, Y. Shirai, H. Ueno***
*Intelligent System Research Division, National Institute of Informatics, 2-1-2 Hitotsubashi,
Chiyoda-ku, Tokyo 101-8430, Japan. (Email: hzaman@grad.nii.ac.jp)
**Department of Computer Controlled Mechanical Systems, Osaka University, Japan.

Abstract - *This paper presents a knowledge-based person-centric human–robot interaction system using facial and hand gestures. In the proposed method, face detection and person identification are first made. With the knowledge of the known user face and hand poses are then classified from the later image frame by the subspace method and the gestures are finally recognized. The rules for interpreting the gestures are selected according to each specific user recognized by the facial image. The user's name and gesture commands are sent to the robot through a Software Platform for Agent and Knowledge Management (SPAK) to implement person-centric human-robot interaction. The effectiveness of this method has been demonstrated by interacting with a humanoid robot Robovie.*

Keywords: Gesture, person-centric human-robot interaction, Subspace method, SPAK.

1 Introduction

The study of human-robot symbiotic systems have been increasing recently considering that a robot will play an important role in the future welfare society. Research in robotics focused on building robots that can be used by ordinary people in their homes, their workplaces, and in public spaces such as hospitals and museums. To realize a symbiotic relationship between human and robot, it is crucial to establish human-robot natural interaction. Ueno [1] presented symbiotic information system and human-robot symbiotic system where human and robot can communicate with each other in human way using speech and gesture. Most gestures are made by hands. But hand gestures have different meaning in different culture. Different users can use the same gesture for activating different actions of a robot. The skin colors of the hand region, hand shapes and hand poses are also different for different person. For realizing reliable gesture-based human-robot interaction person-centric knowledge is the prime factor. There are significant amount of researches on hand, arm and facial gesture recognition to control robot or intelligent machine in recent years. Watanabe *et al* [2] 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. Rigoll *et al* [3] used HMM-based approach for real-time gesture recognition. In that work features are extracted from the differences between two consecutive images and considered that the target image is always in the center of the input images. But practically it is difficult to maintain such condition. Utsumi *et al* [4] 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* [5] detected and tracked face and eye for human robot interaction. 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 mention papers focused on visual processing and did not deal with the knowledge of the different users for gesture interpretation or human robot interaction.

In this paper a knowledge-based person-centric human-robot interaction system using facial and hand gestures is presented. Fig. 1 shows an overall architecture of this system. This system first detects human face using multiple features and recognizes the user using eigenface method [6]. Then using the knowledge of the identified person's profile, face and hand poses are classified as well as gesture are recognized from the later image frames. The person profile keeps the threshold values for chrominance and luminance components of face and hand skin colors and the rules for gesture recognition for each known person. The values of the chrominance and luminance components are defined from statistical analysis of the skin regions while a new user is registered. Face and hand poses are segmented using person-specific skin color information and classified using the subspace method based patterns matching approach. In this system three largest skin like regions are segmented from the input images using person-specific skin color information from the YIQ color space [7,8]. If the combination of three skin-like regions at a particular image frame matches with the predefined gesture of a specific person, then corresponding gesture command is generated. The person's name and gesture name are being sent to SPAK [9] for person-centric human-robot interaction. Using the received gesture and user information, SPAK inference engine processes the facts and activates the corresponding frames to carry out predefined robot action. Gesture commands and robot actions are interpreted in voice so that human can hear which gestures he/she made and which actions are accomplished by the robot [10].

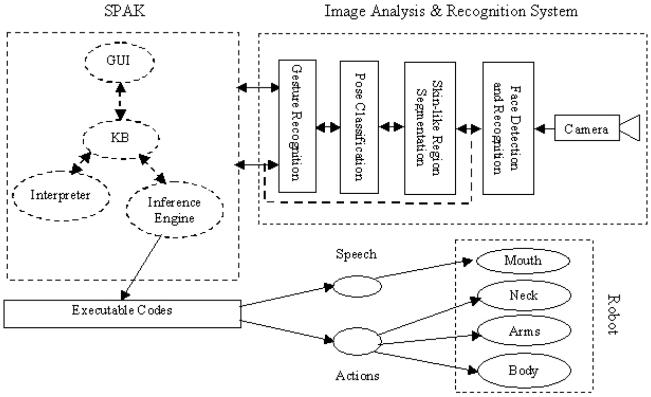


Figure 1. Proposed gesture-based human-robot interaction system architecture

This research has combined computer vision and knowledge-based approaches for person-centric human-robot interaction so that user can define or edit robot behavior according to his desire. User can also define or edit the rules for gesture recognition in the user profile data. The segmented skin regions are more noise free for known person because the probable hand and face poses are segmented using person-centric threshold values for YIQ components. To achieve better accuracy this system uses subspace method or separate eigenspaces for hand and face poses classification instead of normal PCA method. Both static and dynamic gestures are included in this system by tracking the transition of face poses with the classification of static poses. As an application of this method this system has implemented a real-time human robot interaction system using a humanoid robot named Robovie.

This paper is organized as follows. Section 2 briefly describes face detection and person identification methods. Section 3 describes the skin regions segmentation and normalization methods as well as face and hand poses classification method. The gesture recognition method is presented in section 4. Section 5 describes person-centric human-robot interaction using SPAK. Section 6 presents the experimental results and discussions. Section 7 concludes this paper.

2 Person identification

2.1 Face Detection

There are several approaches of face detection, such as knowledge based, facial features invariant, template matching and appearance based methods [11]. This paper has combined template matching and feature invariant approaches for face detection because if only template base method is used then some hand poses near to elliptical shapes may be detected as face. This method uses face template pyramid with different resolutions and orientations. The face templates are moved onto every position of the input image and the matching probability is calculated using Manhattan distance [12]. If the minimal Manhattan distance is less than the predefined threshold value then a search is done for the two eyes on the upper part of the probable face to make sure of the presence of a face [5]. If two eyes are found in any probable face area then the face area is bounded by a square box with the size of the matched template. Fig. 2 depicts the face detection method with example output. 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.

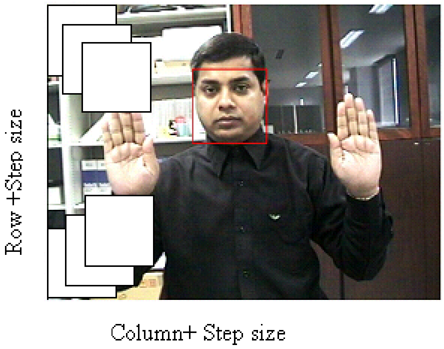


Figure 2. Face detection method

2.2 Person Identification

The detected face is filtered in order to remove noise 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. The face pattern is classified using the eigenface method [6] 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 Fig.3 (1^{st} row). The eigenvectors are calculated from the known person face images for each face class and chosen k-number of eigenvectors corresponding to the highest eigenvalues to form principal components for each class. The minimum Euclidean distance is determined among the weight vectors generated (by projecting them onto the eigenspaces) from the training images and the detected face. 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 [13].

3 Hands and Face Poses Classification

3.1 Skin Region Segmentation and Normalization

Human skin color has been used and proven to be an effective feature in many applications, from face detection to hand tracking. But different people have different skin colors, i.e., chrominance and luminance components are different for different persons. Considering these facts person-specific threshold values for the chrominance and luminance components are used for skin-like regions segmentation. Several color spaces have been utilized to label pixels as skin including RGB, HSV, YCrCb, YIQ, CIE XYZ, CIE LUV etc. However, such skin color models are not effective where the spectrum of the light sources varies significantly. 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. The RGB image taken by the video camera are converted to YIQ color representation system and threshold it by the skin color range of identified person [7,8]. The user profile consists of the threshold values for the chrominance and luminance components of the skin colors of each known person. Probable hands and face regions are isolated from the image with the three largest connected regions of skin-colored pixels. In this system, 8-pixels neighborhood connectivity is employed. 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). After thresholding, the segmented image may be encountered by some holes in the three largest skin-like regions. In order to remove noise and holes, segmented images are filtered by morphological dilation and erosion operations. The dilation operation is used to fill the holes and the erosion operations are subjected to the dilation results to restore the shape. If the person shirt's color is near to skin color then segmentation outputs quality is very poor. If the person wears T-shirt then it needs to separate hand palm from arm. This system considers the person wears 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 [8]. Each segment is scaled to be square image with (60×60) and converted to be gray image. Outputs of the normalization algorithm that look like the training images as shown in Fig. 3.

3.2 Subspace Method for Pose Classification

The main idea of the subspace method is similar to principal component analysis (PCA) method 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 subspaces methods training images are grouped for face and hand poses separately. In subspace method test image is projected on each subspace separately. The approach of face and hand pose classification using subspace method includes the following operations:

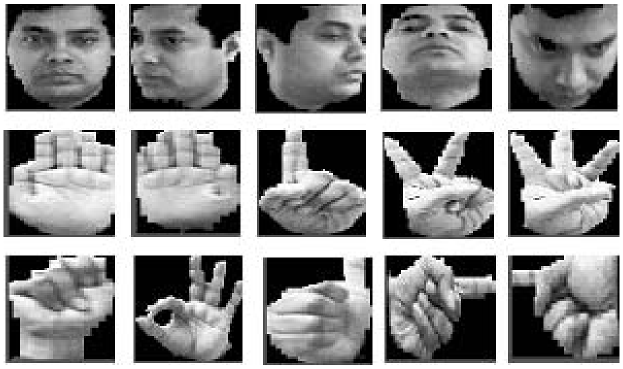


Figure 3 Example of training images

(I) 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 i^th class and j=1,2…. , M. Fig. 3 shows the example training images: frontal face, right directed face, left directed face, up directed face, down 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, point right by index finger are pose P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15 respectively.

(II) For each group, calculate eigenvectors $(u_m^{(i)})$ using Matthew Truk and Alex Petland technique [6] and choose k-number of eigenvectors $(u_k^{(i)})$ corresponding to the highest eigenvalues to form the principal components for that class. These vectors for each group define the subspace of that group.

(III) Calculate corresponding distribution in k-dimensional weight space for the known training images by projecting them onto the subspaces (eigenspaces) of the corresponding group and determined weight vectors $(\Omega_l^{(i)})$, using equations (1) and (2).

$$\omega_k^{(i)} = (u_k^{(i)})^T (s_l^{(i)} - \Phi_i) \tag{1}$$

$$\Omega_l^{(i)} = [\omega_1^{(i)}, \omega_2^{(i)}, ..., \omega_k^{(i)}] \tag{2}$$

Where, average image of i-th class $\Phi_i = 1/M \sum_{n=1}^{M} T_n$ and $s_l^{(i)}(N \times N)$ is l-th known images of i-th class.

(IV) Each segmented region is treated as individual input image and transformed each into eigenimage components and calculated a set of weight vectors $(\Omega^{(i)})$ by projecting the input image onto each of the subspace as equations (1) and (2).

(V) Determine if the image is face pose or other predefined hand pose based on minimum Euclidean distance among the weight vectors using equation (3) and (4),

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

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

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 through experiment, considering optimal separation among the poses.

Table 1: Three segments combination and corresponding gesture (X=absence of predefined hand poses or face poses)

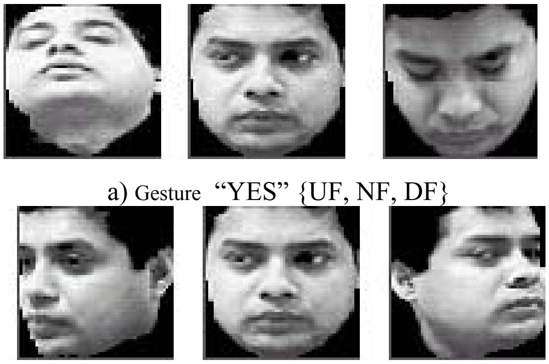
Gesture Components			Gesture names
Face	Left hand palm	Right hand palm	TwoHand
Face	Right hand palm	X	RightHand
Face	Left hand palm	X	LeftHand
Face	Raise index finger	X	One
Face	Form V sign with index and middle fingers	X	Two
Face	Raise index, middle and ring fingers	X	Three
Face	Thumb up	X/Thumb up	ThumbUp
Face	Make circle using thumb and index finger	X	OK
Face	Fist up	X/Fist up	FistUp
Face/X	Point left by index finger	X	PointLeft
Face/X	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

4 Person-Centric Gesture Recognition

The sequence of poses and combination of poses are analyzed for the occurrence of gesture. The rules for recognizing gestures are predefined by the user and it may vary from person to person. To accommodate different user's desires, user profile maintains the person identity, the rules of gestures and gesture commands. If person is unknown then default gesture recognition rules are applied. The system in this research recognizes 13 gestures: 11 static gestures and 2 dynamic facial gestures as listed in Table 1. It is possible to recognize more gestures including new poses and new rules using this system.

4.1 Static Gesture Recognition

The static gestures are recognized using rule-based system with the combination of the pose classification results of the three skin-like regions at a particular time. The user predefines these rules in a person's profile data. For example, if left hand palm, right hand palm and one face are present in the input image then recognizes it as a "TwoHand" gesture. If one face and left hand palm are present in the input image frame then it is recognized as a "LeftHand" gesture. Similarly others static gestures as listed in Table 1 are recognized for a specific person.



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

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

Figure 4 Example dynamic gesture sequences.

4.2 Dynamic Gesture Recognition

Two dynamic facial gestures are recognized in this system considering the transition of the face in a sequence of time steps. If human face shakes left and right then it is defined as "NO" gesture. If human face shakes up and down then it is defined as "YES" gesture. For this purposes this method uses a 3-layers queue (FIFO) that holds the values of 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 first row of Fig. 3 from left to right. For every image frame, face pose is classified using the subspace method. If pose is classified as predefined face pose then it updates the 3-layer queue values. If the classified pose value is same as previous frames then queue values will remain unchanged. From the combination of the 3-layers queue values this method determine the gesture. For example if the queue's values are as {UF, NF, DF} or {DF, NF, UF} pose sets then recognized it as "YES" gesture. Similarly, if the queue's values are as {RF, NF, LF} or {LF, NF, RF} pose sets then recognized it as "NO" gesture. After a specific time period the queue values are refreshed. Fig. 4 shows the example face sequences for dynamic gestures "YES" and "NO".

5 Person-Centric Human-Robot Interaction

The image analysis and recognition units send person identity and gesture command to a knowledge based software platform for decisions making and activating the robot. According to gesture and user identity, the knowledge module generates executable code for predefined robot actions. The robot responds in accordance to gestures by using speech, body actions and its movements. Fig. 5 shows the knowledge hierarchy for person-centric human-robot interaction. This knowledge model is represented using frame-based approaches. The frames are created for users, gesture, robot and robot actions or robot behaviors. User frames include child frames of the known users (P1, P2, .., Pn), gesture frames include all the gestures (G1, G2, .., Gn) as child frames, robot frames include all the robots frames (R1, R2, …, Rn) that are used by the users and robot behavior frames include all the robot actions (A1, A2, …, An) as child frames for a specific robot.

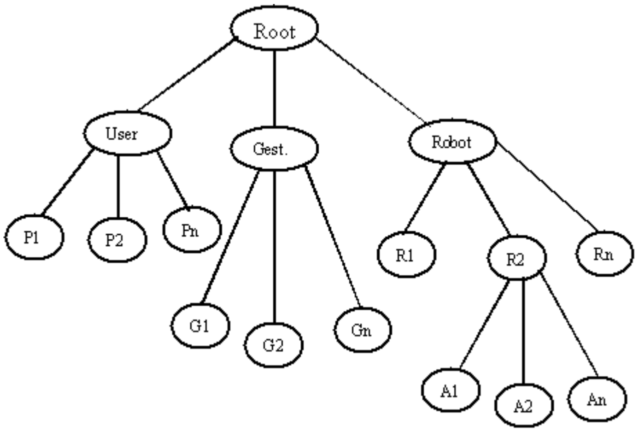
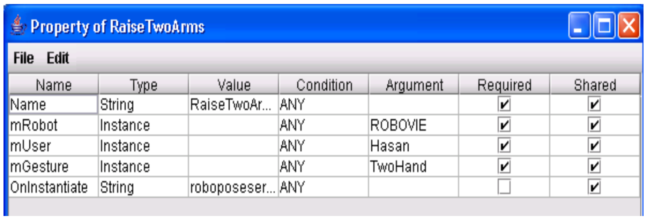


Figure 5 Knowledge model for human robot interaction

This system uses SPAK which consists of a frame based knowledge management system and a set of extensible autonomous software agents representing object inside the environment and supporting human robot interaction and collaborative operation with distributed working environment [9]. SPAK consists of the following major components: GUI interface, Knowledge Base [KB] and Inference engine as shown in Fig.1. SPAK allows TCP/IP-based communication with other software agents in the network and provides knowledge access and manipulation via TCP port. Frame based knowledge is entered into the SPAK system with full slot (attribute) information: conditions, actions. Based on the information from connected agents (e.g. gesture recognition, face recognition output) SPAK inference engine processes facts, instantiates frame instances and carries out the user predefined actions.



Name	Type	Value	Condition	Argument	Required	Shared
Name	String	RaiseTwoAr...	ANY		☑	☑
mRobot	Instance		ANY	ROBOVIE	☑	☑
mUser	Instance		ANY	Hasan	☑	☑
mGesture	Instance		ANY	TwoHand	☑	☑
OnInstantiate	String	roboposeser...	ANY		□	☑

Figure 6 Example frame for the action “Raise Two Arms”

Fig. 6 shows an example of a robot action frame “RaiseTwoArms”. This frame will be activated if the user is “Hasan”, the gesture is “TwoHand” and the selected robot is “Robovie”. The user can define different actions for the same gesture. For example, a user selects the robot Robovie for interaction. The user comes in front of the Robovie eye cameras and it recognizes the person “Hasan” and delivers greeting messages “Hi Hasan, How are you?”. The User “Hasan” raises his Thumb. The gesture recognition module recognizes that the gesture is “ThumbUp” and the face recognition module identifies the person as “Hasan”. In this combination, Robovie replies by speech “You do not look fine, do you want to play now?”. In the case of another user “Cho” uses “ThumbUp” in similar situation, but Robovie replies as “Oh good, Do you want to play now?”. This example expresses that the same gesture is used for different meaning for different persons.

6 Experiments and Discussions

6.1 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 human-robot interaction system using a humanoid robot Robovie developed by ATR [14]. Robovie eye cameras are used for capturing the images. First, the system is trained using the training images for 15 poses (5 face poses and 10 hand poses) of 7 persons. All the training images are 60×60 pixels gray images. The training images consist of 2100 images; 140 images for each pose of 7 person. This system is tested for real time input images as well as static images.

6.2 Results of Recognition

The example visual output of gesture recognition system is shown in Fig. 7(a). It shows the gesture command at the bottom text box corresponding to matched gesture (“Raises Two Hand”). In the case of no match it shows “no matching found” in the bottom text box. Table 2 shows the comparison of precision and recall rate of the subspace method and the standard PCA method for face and hand poses classification. In this comparison 2130 test images of seven persons are used for 15 poses. The threshold value for the classifier is selected so that all the poses are classified. From the results we conclude that precision and recall rate is increases in subspace method and wrong classification rate is decreases. 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 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”. Accuracy of the dynamic gesture recognition also depends on the accuracy of the face pose classification unit.

Table 2: Comparison of subspace method and PCA method

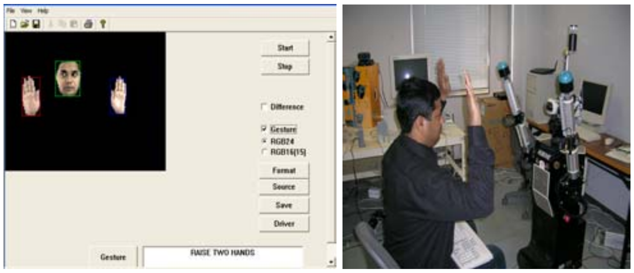
Pose #	Precision (%)		Recall (%)	
	Subspace	PCA	Subspace	PCA
P1	96.21	90.37	97.69	93.84
P2	100	96.59	98.06	91.61
P3	100	93.28	99.28	99.28
P4	97.33	92.30	99.31	97.95
P5	99.21	90.90	98.43	93.75
P6	100	100	94.28	91.42
P7	97.22	96.47	100	97.85
P8	95.17	94.52	98.57	98.57
P9	97.77	97.67	94.28	90
P10	97.81	93.05	95	95
P11	100	100	92.66	87.33
P 12	96.71	96.68	98	97.33
P 13	99.31	100	94.66	93.33
P14	94.89	93.28	97.69	93.84
P15	100	100	100	99.33

The propose face detection method in this paper is robust against background, motion and distance, but this method has a larger computation cost that is the bottleneck for real time human-robot interaction. Three factors directly affect on computation costs: step size, template images dimension and template images number. If step size is 1, the template image dimension is 60×60 and input image dimension is 320×240 then 46800 comparisons are required to sliding one template on the whole image. In similar cases if step size is 2, 3, 4, 5 then numbers of comparisons are 11700, 5220, 2925, and 1872 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 area segmentation and human skin area segmentation. In this work we use human skin area segmentation with reasonable step size to reduce the processing time.

In our previous research [13] we found that the accuracy of frontal face recognition is better than up, down, left and right directed faces. In this system we prefer frontal and a small left-right rotated face for person identification. We have tested this face recognition method for 680 faces of 7 persons, where two are female. The average precision for face recognition is about 93% and recall rate is about 94.08%.

6.3 Implementation of Human-Robot Interaction

The real-time gesture-based human-robot interaction is implemented as an application of this system. The communication link between the robot and the PC has been established through SPAK. Initially, the client PC is connected with the robot server and then the face and gesture recognition program is run in the client PC. As a result of face and gestures recognition program, the client PC sends person name and gesture command to the SPAK. After getting person name and gesture name SPAK inference engine processes facts, instantiates frame instances and activates the corresponding robot action frames. The robot acted according to users predefined actions. This system has considered that gesture command will be effective until robot finishes corresponding action for that gesture.



a) Visual output b) Robot action

Figure 7 Sample output of human-robot interaction

This approach has been implemented on a humanoid robot Robovie for the following scenarios:

User: “Hasan” comes in front of Robovie eyes camera
Robot: “Hi Hasan, How are you?” (speech)
Hasan: uses the gesture “ThumbUp”
Robot: “ You do not look fine, do you want to play now?”
Hasan: uses the gesture “OK”,
Robot: “Oh good”
Hasan: uses the gesture “TwoHand”
Robot: imitate user’s gesture (“Raise Two Arms”)

Similarly, Robovie imitates for other gestures. The other actions that Robovie imitates are, “Raise Left Arm”, “Raise Right Arm”, “Move Neck Left-Right or Right-Left”, “Move Neck Up-Down or Down-Up” corresponding to the gestures “LeftHand”, “RightHand”, “NO” and “YES” respectively for the user “Hasan”.

Hasan: uses the gesture “FistUp” (stop the action)
Robot: Bye-bye (speech).

For another user “Cho”,
User: “**Cho**” comes in front of Robovie eyes camera
Robot: “Hi Cho, How are you?”
Cho: uses the gesture “ThumbUp”
Robot: “ Oh, good, do you want to play now?”
Cho: uses the gesture “Two (V sign)”,
Robot: “Thanks!”
Cho: uses the gesture “TwoHand”
Robot: imitate user’s gesture (“Raise Two Arms”)

Similarly, Robovie imitates for other gestures. The other actions of Robovie are, “Raise Left Arm” and “Raise Right Arm”, corresponding to the gesture “LeftHand” and “RightHand”.

Hasan: uses the gesture “NO” (shakes face left-right)
Robot: Bye-bye (speech)

The above scenario shows 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 using Robovie and can design corresponding knowledge frame using SPAK to implement their desired actions.

7 Conclusions

This paper describes a knowledge-based person-centric human-robot interaction system using facial and hand gestures. Human skin-color (luminance and chrominance components) differ from person to person so person centric threshold values for YIQ components is very useful for skin region segmentation. This system uses separate eigenspaces for face and hand poses classification that is more reliable than normal PCA based method. In addition, with the gesture recognition this system is also capable to identify persons. By integrating with knowledge-based software platform in this system, gestures-based person-centric human-robot interaction has also been successfully implemented using a robot Robovie. In this system the user can define or update the rules for gesture recognition and the robot behaviors corresponding to his 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. Our next approach is to make the system more robust and to recognize more static and dynamic gestures for interaction with different robots such as AIBO, Robovie, SCOUT, etc. The 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.

References

[1] Haruki Ueno, “A Knowledge-Based Information Modeling for Autonomous Humanoid Service Robot”, IEICE Trans. On Information & Systems, Vol. E85-D, No. 4, pp. 657-665, 2002.

[2] Takahiro Watanabe, Masahiko Yachida, “Real-time Gesture Recognition Using Eigenspace from Multi-Input Image sequences”, System and Computers in Japan, Vol. J81-D-II, pp. 810-821, 1998.

[3] Gerhard Rigoll, Andreas Kosmala, Stefan Eickeler, “High Performance Real-Time Gesture Recognition Using Hidden Markov Models”, In proc. Gesture and Sign Language in Human Computer Interaction, International Gesture Workshop, Germany, pp. 69-80, 1997.

[4] Akira Utsumi, Nobuji Tetsutani and Seiji Igi, “Hand Detection and Tracking using Pixel Value Distribution Model for Multiple-Camera-Based Gesture Interactions”, Proc. of the IEEE workshop on knowledge Media Networking (KMN’02), pp. 31-36, 2002.

[5] 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.

[6] Matthew Turk and Alex Pentland “Eigenface for Recognition” Journal of Cognitive Neuroscience, Vol. 3, No.1, pp. 71-86, 1991.

[7] Md. Al-Amin Bhuiyan, Vuthichai Ampornaramveth, Shin-yo Muto, Haruki Ueno “Face Detection and Facial Feature Localization for Human-machine Interface”, NII Journal., No. 5, pp. 25-39, 2003.

[8] Md. Hasanuzzaman, V. Ampornaramveth, T. Zhang, M.A. Bhuiyan, Y. Shirai, H. Ueno, "Real-time Vision-based Gesture Recognition for Human-Robot Interaction", Proc. of the IEEE Int. conf. on Robotics and Biomimetics (ROBIO), China, pp. 379-384, 2004.

[9] Vuthichai Ampornaramveth, Haruki Ueno, “Software Platform for Symbiotic Operations of Human and Networked Robots”, NII Journal, Vol.3, pp 73-81, 2001.

[10] “The Festival Speech Synthesis System developed by CSTR” http://www.cstr.ed.ac.uk/project/festival.

[11] Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja, “Detectin Faces in Images: A survey”, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 24, No. 1, pp. 34-58, January 2002.

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

[13] Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, M.A. Bhuiyan, Y. Shirai, H. Ueno, "Face and Gesture Recognition Using Subspace Method for Human-Robot Interaction", Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia, LNCS (Springer-Verlag Berlin Heidelberg) Vol. 3331(1), pp. 369-376, Tokyo, Japan, 2004.

[14] Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai and Ryohei Nakatsu, “Development and Evaluation of an Interactive Humanoid Robot "Robovie",” IEEE International Conference on Robotics and Automation (ICRA 2002), pp.1848-1855, 2002.