Gesture Recognition for Human-Robot Interaction Through a Knowledge Based Software Platform

M. Hasanuzzaman[1], Tao Zhang[1], V. Ampornaramveth[1], M.A. Bhuiyan[2], Yoshiaki Shirai[3], and H. Ueno[1]

[1] Intelligent System Research Division, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.
hzamancsdu@yahoo.com
[2] Jahangirnagor University, Dhaka-1342, Bangladesh.
[3] Department of Computer Controlled Mechanical Systems, Osaka University, Suita, 565-0871 Japan.

Abstract. The task of real-time hand gesture recognition is extremely challenging due to a number of DOFs of hand pose and motion. However, for human-robot interaction in natural ways, gesture can provide a powerful interface tool for commanding a robot to perform a specific task. This paper presents a vision-based real-time gesture recognition system by segmenting the three largest skin color components and template-matching techniques with multiple features. Gesture commands are generated whenever the combinations of three skin-like regions at a particular image match with the predefined gestures. These gesture commands are sent to robots through a knowledge based software platform for human-robot interaction. The effectiveness of our method has been demonstrated by interacting with an entertainment robot named AIBO.

1 Introduction

Research on humanoid robots has been increasing rapidly for the last decade. Recent advances in computer science and robotics make robots more applicable to our daily life. Robots and human will co-exit with sharing and co-operating tasks according to Symbiotic Information System (SIS) concept [1]. However, most robots so far still lack of the ability to interact with user in natural ways and work independently. Recent technologies for multi-modal communication can provide various communication channels like voice and gestures. Using gesture to interact with a robot provides a natural and powerful interface that can be used to command a robot to perform tasks or to control its operations. Considering the facts it is necessary to develop efficient and real time gesture recognition system to perform more human like interaction between human and robot. Two approaches are commonly used to interpret gestures for human machine interaction. One is gloved based approach [2] that requires wearing of cumbersome contact devices and generally carrying a load of cables that connect the device to a computer. 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. Gesture recogni-

A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3211, pp. 530–537, 2004.

tion based on vision technology has been emerging with the rapid development of computer hardware of vision system in recent years and in future it will dominate in both Human-Computer and Human-Robot interactions. For gesture interpretation system gestures modeling is the first step that mainly depends on the intended application of those gestures. Gesture modeling can follow appearance based or model based approach. Model based approach is very hard to implement in real time because they usually use very complicated algorithms to extract accurate joint angles. An appearance-based algorithm is a strong tool for object recognition. Here, a variety of object appearances are stored as a statistical model and used in the recognition task. The gestures are modeled by relating the appearance of any gesture to the appearance of the set of predefined template gestures.

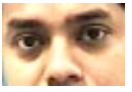
Watanabe et. al. [3] used maskable template based on minimum distance between template and partial block of an input image for gesture recognition. Hongo et. al. [4] has developed a system that can track multiple faces and hands by using multiple cameras to focus on face and gesture recognition. Utsumi et. al. [5] detected hand using hand shape model and tracked using extracted color and motion. They also propose multiple cameras for data acquisition to reduce the occlusion problem. But in this process there incurs complexity in computations. However all this paper [3-17] did not consider face and two hands gestures at the same time that we have considered here.

In this paper we present a simple and faster method for recognizing gestures with skin-color segmentation and multiple features-based template matching techniques. In this method three larger skin like regions are segmented from the input images by skin color segmentation technique considering that face and two hands may be present in the input images at the same time. Segmented blocks are filtered and normalized and then compared with template images for finding best match. For template matching we have used the combination of two features: correlation coefficient and minimum (Manhattan distance) distance qualifier. If the combination of three skin-like regions at a particular frame matches with predefined gesture then corresponding gesture command is generated. In this experiment we have recognized eight gestures as listed in Table 1. Gesture commands are being sent to robots through a knowledge-based software platform and their actions are being accomplished according to users predefined action for that gesture. A method has also been developed to detect left hand and right hand relative to face position, as well as, to detect the face and locate its position. We have prepared template images with different illuminations to adapt our system with illumination variation. In this method we have also considered slightly rotated face for both left and right side to make face detection algorithm a more pose invariant. As an application of our method, we have implemented a real-time human robot interaction system using a robot named AIBO.

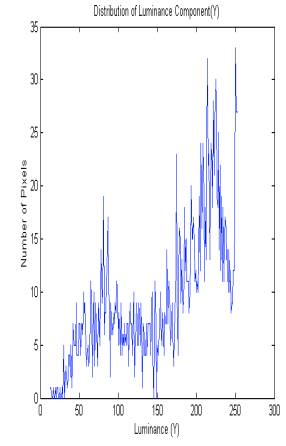
The remainder of this paper is organized as follows. In section 2, we briefly describe skin-like regions segmentation, filtering as well as normalization. Section 3 focuses on multiple features based template-matching techniques for face and hand poses detection and gestures recognition method. Section 4 presents our experimental results and discussion. Section 5 concludes this paper.

2 Skin Color Segmentation, Filtering, and Normalization

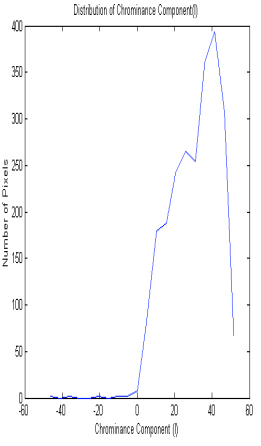
This section introduces a color segmentation based approach for determining skin parts (mainly face and hands) from color images. This system uses video camera for data acquisition and from color input image isolate three larger skin-like regions by using skin-color information. HSV or YIQ color model is used for skin color segmentation, since color footprint is more distinguishable and less sensitive to illumination changes in the hue saturation space than the standard RGB color space. We use YIQ (Y is luminance of the color, I, Q are chrominance of the color) color representation system since it is typically used in video coding and provides an effective use of chrominance information for modeling the human skin color.



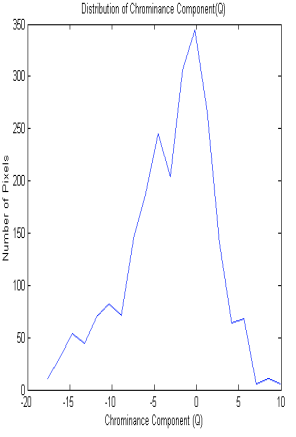
(a)



(b)



(c)



(d)

Fig. 1. a) Sample skin region, b), c) and d). Y, I, Q distributions respectively

The RGB image taken by video camera are converted to YIQ color space and threshold it by the skin color range [6]. Fig. 1., shows sample skin region and corresponding Y, I, Q distributions for every pixels. Chrominance component I, play an important role to distinguish skin like regions from non-skin regions, because it is always positive for skin regions and negative for non-skin region. Locations of the probable hands and face are determined from the image with three larger connected regions of skin-colored pixels. In this experiment, 8-pixels neighborhood connectivity is employed. In order to remove the false regions from the isolated blocks, smaller connected regions are assigned by the values of black-color. Noise and holes are filtered by morphological dilation and erosion operations and normalization is done to con-

vert the segmented images to gray images and resized the image as template images size. Sample output of the skin-color segmentation is shown in Fig. 2.

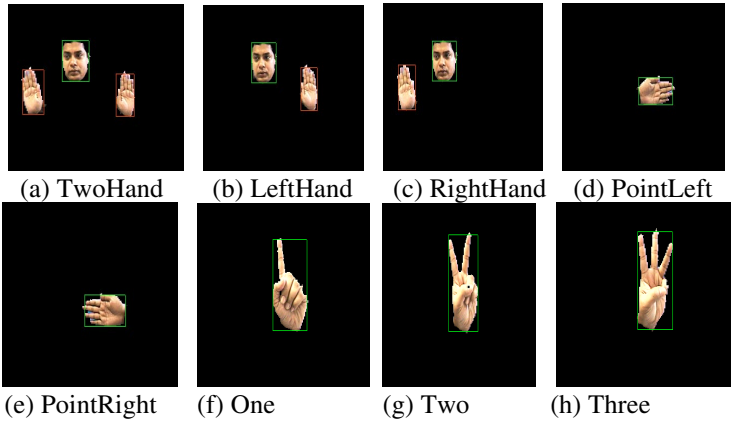


Fig. 2. Sample output of Skin color Segmentation and filtering

3 Face and Hand Poses Detection for Gesture Recognition

3.1 Hand Poses and Face Detection Using Multiple Features Based Template Matching

As template-matching approach is a more natural approach of pattern recognition, we use it to recognize gesture from an unknown input image. First we have prepared noise free version of faces and hands as templates as shown in Fig. 3. To support small rotation we included some slightly rotated images within our template images. To adapt illumination variation we have formed templates in different illumination. For template matching we have considered two features: one is maximum correlation coefficient and another is minimum distance classifier (Manhattan distance) between two same size images. Correlation coefficient is calculated using following equation,

$$\alpha_t = M_t / P_t \quad (0<\alpha_t \quad 1) \tag{1}$$

where M_t is total number of match pixels (white pixels with white pixels and black pixels with black pixels) with t-th template, P_t is number of total pixels in t-th template and t is a positive number. For exact matching α_t is 1, but for practical environment we have chosen threshold value for α_t through experiment for optimal matching. Minimum distance can be calculated by using following equation,

$$\delta_t = \sum_{1}^{x \times y} | I - G | \tag{2}$$

where, I(x, y) is the input image and $G_1(x,y)$, $G_2(x,y)$,----------,$G_t(x,y)$ are template images. There is more than one way to define $_t$ corresponding to different ways of measuring distance. Two of the most common are: Euclidean metric and Manhattan metric. In our experiment we have used Manhattan metric. For exact matching $_t$ is 0 (zero) but for practical purpose we have used a threshold value through experiment for finding optimal matching.

Fig. 3. Example template images

We have combined output of these two matching methods to make our system more accurate to recognize gestures. In our method we have grouped template images into eight classes (each for different hand poses), for example face class (C_1) and left hand class (C_2) and right hand class (C_3), etc. If max $\{\alpha_t\}>th_1$ is true then corresponding class (C_α) is identified and if min $\{ _t\}<th_2$ is true then corresponding class (C) is identified, where th_1 and th_2 are thresholds for correlation coefficient and minimum distance qualifier respectively. If both methods identified the same class then corresponding class is detected, otherwise ignored.

3.2 Gesture Recognition

Similar calculation is done for all three segments simultaneously for a particular image frame and with the combinations of three segments at a particular time form a gesture. Gestures are recognized using rule-based system according to Table 1. For examples: if two hands and one face are present in the input image then recognized it as "TwoHand" gesture. If one face and one hand are present in the input image frame then recognize it as either left hand or right hand respective to face position using following equation,

$$\varepsilon = f_x - h_x \tag{3}$$

where, f_x is the x-coordinate of the centre position of face segment and h_x is the x-coordinate of the centre position of hand segment (when one hand is present). If the distance is negative then it is detected as "RightHand" gesture and if it is positive then it is detected as "LeftHand" gesture. If pointing left hand present in the input images then it is recognized as "PointLeft" gesture and similarly others gesture are recognized. According to gestures recognition corresponding gesture command is generated and sent to interact with robot through a knowledge base software platform (SPAK) [7]. SPAK (Software Platform for Agent and Knowledge Management) consists of a frame based knowledge management system and a set of extensible autonomous software agents representing object inside the environment and support human

robot interaction and collaborative operation with distributed working environment. Our approach has been implemented on a robot named AIBO.

Table 1. Input segment combinations and corresponding gestures.

Three input segment combinations			Gesture
Face	Left hand palm	Right hand palm	Twohand
Face	Right hand palm	X	Righthand
Face	Left hand palm	X	LeftHand
Face/(no face)	Point left hand	X	PointLeft
Face/(no face)	Point right hand	X	PointRight
Face/(no face)	Index finger raise	X	One
Face/(no face)	Form V sign with index and middle fingers	X	Two
Face/(no face)	Index, middle and ring fingers raise	X	Three

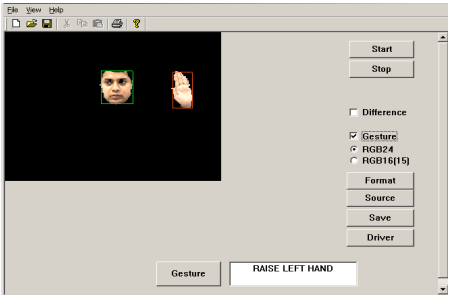
[X=Absence of predefined hand or face poses]

4 Experimental Results and Discussion

This section describes experimental procedures, as well as experimental results of the gestures recognition system and human-robot interaction system by gesture. This system uses a standard video camera for image data acquisition. Each capture image is digitized into a matrix of 320×240 pixels with 24-bit color. First we prepare pure templates for face and different hand poses. All the templates are 60×60 pixels gray images. Template images are consisted a total of 480 frontal images of faces, different hands poses of different persons related to gestures. Fig. 3 shows the example template images. We have tested our system for real time input images. The sample output of our gesture recognition system is shown in Fig. 4(a). This shows gesture command at the bottom text box corresponding to matched gesture, in case of no match it shows “no matching found”.

In this part we have explained a real time human-robot interaction system using recognization gestures. We have implemented this application by off-board configuration that’s means gesture recognition program run in client PC, not in robot. We have considered robot as a server and our PC as a client. Communication link has been established through SPAK. Initially, we connected the client PC with SPAK server and then gestures recognition program was run in the client PC. As a result of gestures recognition program client PC has sent gesture commands to SPAK. In SPAK we have designed frames corresponding to each gesture. After getting gesture command related frame is activated and robot acted according to users predefined actions. Fig. 4(b) shows our experiment setup for human-robot interaction using gestures. The actions of AIBO are: STAND, WALK FORWARD, WALK BACKWARD, TURN LEFT, TURN RIGHT, KICK (by right leg), SIT and LIE in accordance with gesture

"One", "Two", "Three", "PointLeft", "PointRight" "RightHand" "TwoHand", and "LeftHand" respectively. We have considered for human-robot interaction that gesture command will be effective until robot finished corresponding action for that gesture.



(a) Output for "LeftHand" gesture.

(b) AIBO STAND-UP for "One" gesture.

Fig. 4. Sample output of gesture based human-robot interaction system.

5 Conclusions

This paper describes a real-time hand gesture recognition system using skin color segmentation and multiple features based template-matching techniques. For the matching algorithm we have used combinations of minimum distance qualifier (Manhattan distance) and correlation coefficient based matching approaches, that's why it is more robust than any single feature based template-matching techniques. We have compared only precisely segmented images with templates, and used a group of template for a single pose, that's why it is more reliable and it has less computational cost. One of the major constrain of this system is that the background should be non-skin color substrate. If we used infrared camera then it is possible to overcome this problem just by a minor modification of our segmentation technique, other modules will remain the same.

We have also successfully implemented gestures based human-robot interactive system using a robot named AIBO. A particular user may assign distinct command to specific hand gesture and thus control various intelligent robots using hand gestures. The significant issues in gesture recognition for our method are the simplification of the algorithm and reduction of processing time in issuing commands for the robot. Our next approach is to make the detection system more robust and to recognize more facial and hand gestures for interaction with different robots such as AIBO, ROBOVIE, SCOUT, MELFA, etc. Our ultimate goal is to establish a symbiotic society for all of the distributed autonomous intelligent components so that, they share their resources and work cooperatively with human beings.

References

1. Vuthichai Ampornaramveth, Haruki Ueno, "Concepts of Symbiotic Information System and Its Application to Robotics", proceedings of 11th European-Japanese Conference on Information Modeling and Knowledge Bases, 2001, pp. 394-407.
2. Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, " Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review" IEEE PAMI, 1997, Vol. 19, No. 7, pp. 677-695.
3. Takahiro Watanabe, Chil-Woo Lee, Akitoshi Tsukamoto, and Masahiko Yachida, "Real-Time Gesture Recognition Using Maskable Template Model", Proc. of the International Conference on Multimedia Computing and Systems (ICMCS'96), 1996, pp. 341-348.
4. Hitoshi Hongo, Mitsunori Ohya, Mamoru Yasumoto, Yoshinori Niwa, and Kazuhiko Yamamoto, "Focus of Attention for Face and Hand Gesture Recognition Using Multiple Cameras", AFGR00, IEEE, pp. 156-161.
5. 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), 2002, pp. 31-36.
6. Md. Al-Amin BHUIYAN, Vuthichai AMPORNARAMVETH, Shin-yo MUTO, Haruki UENO "Face Detection and Facial Feature Localization for Human-machine Interface", NII Journal. March-2003, No. 5, pp. 25-39.
7. Vuthichai Ampornaramveth, Haruki Ueno, "Software Platform for Symbiotic Operations of Human and Networked Robots", NII Journal, Vol.3,pp 73-81,2001
8. Matthew Turk and Alex Pentland " Eigenface for Recognition" Journal of Cognitive Neuroscience, 1991, Vol. 3, No.1, pp. 71-86.
9. Yu Huang, Thomas S. Huang, Heinrich Niemann, "Two-Hand Gesture Tracking Incorporating Template Warping With Static Segmentation", AFGR'02, IEEE, 2002. pp. 260-265.
10. M.A. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno, " ON TRACKING OF EYE FOR HUMAN-ROBOT INTERFACE", International Journal of Robotics and Automation, 2004, Vol. 19, No. 1, pp. 42-54.
11. Hyeon-Kyu Lee and Jin H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition" 1999, Vol. 21, pp. 961-973.
12. Nobutaka Shimada, Yoshiaki Shirai, "3-D Hand Pose Estimation and Shape Model Refinement from a Monocular Image Sequence", Proc. of VSMM'96 in GIFU, pp.23-428
13. Radek Grzeszczuk Gray Bradski Michael H chu Jean-Yves Bouguet, "Stereo Based Gesture Recognition Invariant to 3D pose and lighting", CVPR'00, IEEE, 2000, pp. 1826-1833.
14. Yunato Cui and John J. Weng, " Hand Segmentation Using Learning-Based prediction and verification for hand Sign Recognition, Proc. of the Conference on Computer Vision and pattern Recognition (CVPR'96), IEEE, 1996, pp. 88-93.
15. Yoichi Sato, Yoshinori Kobayashi, Hideki Koike "Fast Tracking of hands and Fingertips in Infrared Images for Augmented Desk Interface", AFGR'00, IEEE, 2000, pp. 462-467.
16. Charles J. Cohen, Glenn Beach, Gene Foulk. "A Basic Hand Gesture Control System for PC Applications" Proc. of the 30th Applied Imagery Pattern Recognition Workshop (AIPR'01), IEEE, 2001, pp. 74-79
17. Dong, Guo, Yonghua Yan and M. Xie, "Vision-Based Hand Gesture Recognition for Human-Vehicle Interaction", Proc. of the International conference on Control, Automation and Computer Vision, 1998, Vol. 1, pp. 151-155.