Efficient Gesture Interpretation for Gesture-based Human-Service Robot Interaction

D. Guo, X. M. Yin, Y. Jin and M. Xie
School of Mechanical and Production Engineering
Nanyang Technological University
Nanyang Avenue, SINGAPORE 639798
mdguo@ntu.edu.sg, *mmxie@ntu.edu.sg*

Abstract

Service robots are designed to perform useful services for human-like activities in dynamic and unstructured environments, they will be of great technological and economical importance in the near future. The interactive ability of service robot with human is one of its essential characteristics to implement service tasks by cooperating with people efficiently. In this paper, a vision-based interactive model is presented, it attempts to build a compact and intuitive interaction agent among service robot and common users based on gesture interpretation of human body. The 3-D model of human upper body is proposed for stereo measurement of body parts. Human gestures are estimated by the spatial position of upper body parts. A neural model of human body is generated for gesture segmentation in the training procedure of RCE neural network, it is capable of delineating the color distribution of arbitrary human body's appearance in color space. The 3-D positions of body parts are acquired by binocular stereo measurement in the segmented area, attentive regions are defined to search the positions of arm joints in the view of active vision. The joint angles of arms are estimated to determine the gesture of human body.

1. Introduction

There is evident that robotics is reaching its maturity in its original field of manufacturing automation and is rapidly evolving toward service industries because of the increasing importance of service sectors for economic growth [1]. Service robots are distinguished from industrial robots for their service tasks, which are related to the functions of maintenance, transport, or manipulation etc. It is the common understanding that service robots are programmable, sensor-based, free moving appliances that fully or semi-automatically perform useful services to human or machines [2]. They are designed to realize a wide range of services, such as assistance for disables, goods transport, waste disposal, security, housekeeping and entertainment etc.

The variety of possible applications and environments dictates many different design solutions of service robot, humanoid service robot is one of the attractive design patterns that gains more attentions in service robot design [3,4,5,6].

1.1 Humanoid Service Robot JINGANG

Service robots are designed to work in the environment where people co-exist. It is a prerequisite that their layouts should meet such an environmental requirement. The design of humanoid service robot is enlightened by the view of anthropomorphic design, which takes human as the reference prototype with the following application advantages:

1. Environmental Adaptability

The indoor environments are built to take into account the spatial requirements of human (e.g. width and height of doors etc.). Humanoid service robots are designed to possess the appearance similar to human, they are well adapted to perform the service tasks without the need of environment modification.

2. Behavior Representation

The humanoid appearance of service robot has the advantage of imitating human-like behaviors, it is convenient to implement service task planning by behavior representation.

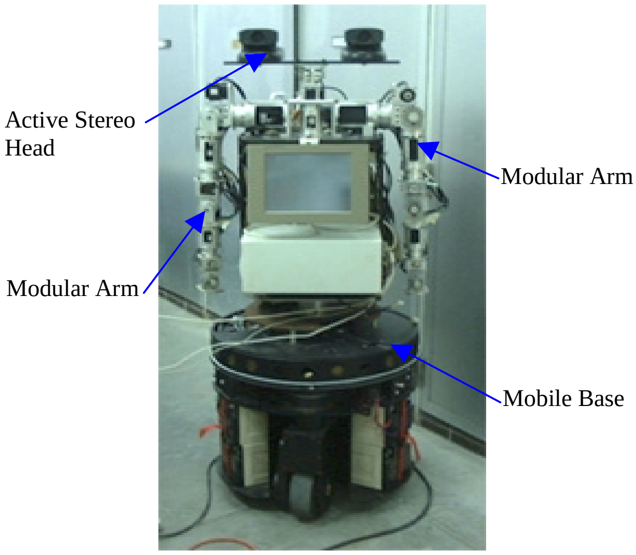

Figure 1 JINGANG Humanoid Service Robot

3. Human-Robot Interaction

Service robots are required to communicate with people for the purpose of efficient human-robot cooperation. Humanoid service robots are able to demonstrate human-like behaviors that make human to interact with robot in a more natural way.

Humanoid service robot is expected to implement diverse service tasks in unknown environments, it is essential to integrate autonomous navigation and dexterous manipulation into the system. To explore its optimal design, a humanoid service robot JINGANG is developed for performing general service tasks in unstructured environments [7]. It mainly consists of two modular arms, omnidirectional mobile base, and active stereo vision head (Figure 1).

1.2 Human-Service Robot Interaction

The development of humanoid service robot brings up many challenging research issues. Human-robot symbiosis and relatively unstructured environment in which service robots must operate are key differences between service robots and industrial robots [2]. Friendly and cooperative human-robot interface is critical for the development of service robot [8]. The need of efficient human-service robot interaction has been recognized in recent years [2]. It is motivating the research of innovative human-robot interfaces that facilitate the interaction between ordinary users and service robots.

Humanoid service robot is designed for public services, it is operated by public users who even might not be able to operate the computer keyboard. It is necessary to implement human-service robot interaction in the manner of human's natural communication preferences (e.g. speech, gesture). Moreover, efficient human-robot interaction can greatly enhance the operation safety, which is main concern for public to accept the existence and usage of humanoid service robots.

Speech and gesture are two natural capabilities that human highly relies on for daily activities, the latter uses hand or body gestures to efficiently convey the ideas that are more easily expressed with actions than words in the noisy environment.

1.3 Gesture-based Interaction

Human gestures are formed by hands and upper limbs, which are dexterous parts with abundant joints in human body. Gesture-based interactions are expected to guide the motion of service robot by using the spatial frame of human gestures. Human can control the service robot where to go or look, when to move, speed up or stop based on different human gestures. Furthermore, service robot can easily acquire the geometrical information of target objects with the help of human gestures. For example, human can tell service robot which object need to be grasped by pointing to it, or guide the service robot to move the spot human points to.

Gesture-based interaction is a process of real-time interaction in which vision system and robotic control are involved and combined, it extends the capability of robotic perception and enhances the intelligent level of service robot.

1.4 Related Works

Gesture-based interaction was firstly proposed by M. W. Krueger as a new form of human-computer interaction in the middle of the seventies [9]. It has become an important research issue with the massive influx of computers in our society, a wide spectrum of available techniques on gesture-based interaction are proposed, which are based on either auxiliary devices or computer vision [10].

Vision-based interaction is gaining more interest with the advantages of intuition, device independence and non-contact, it mainly depends on temporal modeling or spatial modeling of gestures. Important differences in spatial modeling arise depending on whether a 3D model of human hand or an image appearance model of human hand is used to build the model of gestures [11]. 3D gesture modeling aims to build the three-dimensional geometry model of gestures elaborately, it is computational cost and difficult to implement in real-time. Appearance gesture modeling works well under constrained situation, but lacks of generality for natural human computer interaction.

In the case of human-service robot interaction, simply gesture recognition may not be sufficient for service robot to implement diverse service tasks, it needs to be connected to the action of service robot in which gesture recognition of body parts is necessary to be involved for more efficient interactions. R. E. Kahn *et al* [12] developed their PERSEUS gesture recognition system to perform the pointing task, a variety of vision techniques are applied for gesture recognition (e.g. motion, color, edge detection). PERSEUS system is able to recognize the people when they point and find the object people point to, the positions of head and hand are used to determine which area people point to. This system requires static background and relies on off-board computation. S. Waldherr [13] proposed a template-based gesture interface for human-robot interaction, robot is instructed by easy-to-perform arm gestures. An interactive clean-up task is realized by human-robot interaction, a person guides robot to specific location that needs to be cleaned, robot picks up trash and delivers to the nearest trash-bin. The proposed method uses color-based tracking algorithm to follow a person, it suffers the difficult to deal with multi-colored objects. D. Kortenkamp *et al* [14] built a real-time three-dimensional gesture recognition system, a coarse 3D model of human is used to guide stereo measurements of body parts in the view of active vision. The limitation of system is that it can only track one arm at a time.

1.5 Approach Overview

Vision-based gesture interpretation is an effective way for human-service robot interaction. This paper presents a gesture-based interactive model that is applied to instruct service robot JINGANG through human body gestures. The 3D modeling of human body is firstly built for stereo measurement of human body, human gestures are defined by the spatial position of upper body. The color model of human body is further built by using the training procedure of RCE neural network, human body is segmented in RCE running procedure based on training color model. In segmented areas, attentive regions are defined to identify the positions of arm joint by means of binocular stereo measurement. Human gestures are classified by estimating the joint angles of human arms.

2. Spatial Modeling of Human Gestures

Gesture interpretation requires the geometry modeling of human body, which has been extensively studied and used in computer graphics, animation and virtual reality [15]. The model of human body is built to represent the spatial layout of body parts and connectivity of joints, human gestures are determined by sampling the orientations of body parts in the model.

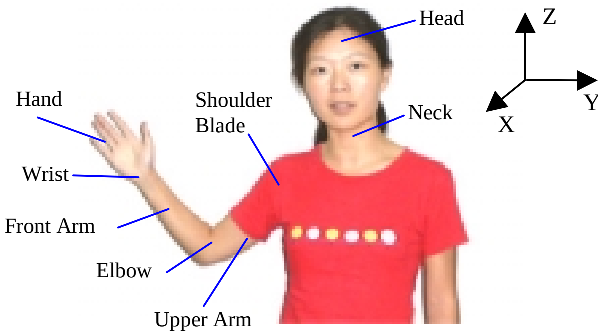

Figure 2 Anatomy of Human Upper Body

2.1 Anatomy of Upper Body

The upper body of human consists of head, torso, upper arms, front arms and hands that are connected by the joints of neck, shoulder blades, elbows and wrists (Figure 2). The motions of joints are similar to the motions of revolved joint, but their ranges are restricted by the anatomy of upper body. D. R. Houy [16] summarized the ranges of joint motions by sampling 100 college students. Table 1 lists the motion ranges of some joints that are essential for gesture interpretation, degree ranges of left arm rotations are indicated under the reference coordinate in Figure 2.

Table 1 Degree Ranges of Body Parts Rotations

Motion	Range of Rotation (Degree)	Datum Axis
Head to Torso	Y: ± 38	Z
Upper Arm to Shoulder	X: +75, -70 Z: ±45	-Y -Y
Front Arm to Upper Arm	Upper Arm: +110, -110	Upper Arm

2.2 3-D Modeling of Human Body

Gesture interpretation aims to classify the orientations of human arms relative to human body, upper arm and front arm are mainly concerned in gesture interpretation. The geometric model of human boy is expected to imitate the skeleton of human body and its joint motions with comparatively low computation cost. A geometric model with hierarchical architecture is used for 3-D modeling of human body in Figure 3, it is able to represent the body parts in a logical way. The model consists of head, torso, upper arm, and front arm with the revolved links relative to them. The motions of revolved links are restricted by the criteria in Table 1.

2.3 Estimation of Upper Body Gestures

In gesture-based interaction, service robot receives the control command based on the interpretation of different gestures. The meaning of different upper body gestures has been studied and defined in the literature. S. Waldherr [13] selected three kinds of gestures (stop, follow and pointing) to guide the operation of service robot, D. Kortenkamp [14] proposed six basic gestures (pointing, thumbing, relaxing, raised, arched and halt) for human-robot interaction. However, a definition of upper body gesture is not useful if it has no meaning to implement service tasks. As a typical gesture of upper body, pointing gesture is commonly used to guide others to find the object along with the pointing direction in human communication, it is also useful to implement service tasks like pick-up or floor cleaning in human-service robot interaction.

In the model of upper body, pointing gesture has the property that upper arm is approximately co-linear with front arm, it can be defined by estimating the angles of should and elbow links (see Figure 3). The orientation of human arm is regarded as pointing gesture if $40° < \alpha < 135°$ and $135° < \beta < 225°$.

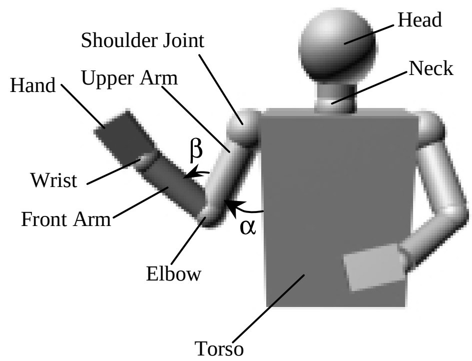

Figure 3 3-D Model of Human Upper Body

3. Color-based Gesture Segmentation

Human gesture segmentation is a procedure of separating human upper body from complex image background, it is the first important step for gesture-based interaction. The colors of human upper body are important perceptual features that offer more robust information under partial occlusion, rotation, scaling and resolution changes. The colors of human skin and outer clothing have their specific distributions in color space, they can be clustered to form a feature space for gesture segmentation.

3.1 Color Distribution of Objects

The visible colors are represented in 3D color space, RGB, L*a*b* and HIS are three color spaces that are commonly used in color vision. In this paper, color-based gesture segmentation is implemented by color clustering in L*a*b* color space, because L*a*b* color space is a uniform color space that has low computation cost in color space conversion. It is observed that color objects of human skin and outer clothing have the following distribution properties:

1. The colors of objects distribute in small regions of color space.
2. The colors of objects do not randomly fall into some regions, but to form clusters at specific points.

Figure 4 shows the color distribution of human skin in Figure 2, it has irregular appearances in L*a*b* color space.

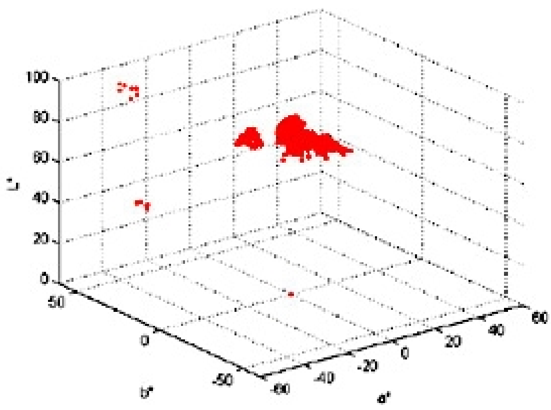

Figure 4 Color Distribution of Human Skin

3.2 Color Modeling by Learning

Common segmentation algorithms are difficult to segment the color objects with irregular appearances due to the problem of proper threshold selection. A cluster- oriented segmentation algorithm is proposed to segment the gesture by color prototypes of objects derived from experience learning. Color prototypes of objects refer to the abstract representations of object colors, they form different spherical influence fields that are able to accurately bound the distribution regions of object colors in color space.

3.3 Gesture Segmentation

RCE neural network is capable of performing adaptive pattern classification by applying the supervised training algorithm to generate prototypes and define values of network connection [17]. Figure 5 shows the network architecture used for gesture segmentation. In network training, object colors are extracted by estimating the density of color prototypes, they are stored in the prototype layer of RCE neural network. With various color prototypes built in learning procedure, RCE neural network generates the segmentation results in fast response model.

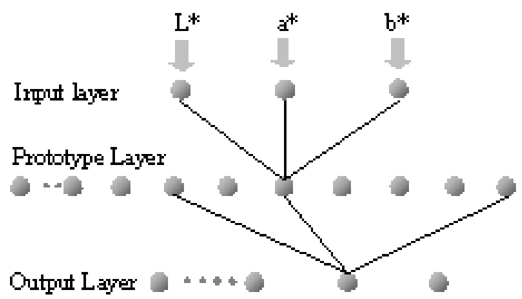

Figure 5 The architecture of RCE neural network

4. Active Perception of Human Gesture

D. H. Ballard [18] promoted the view of active vision to reduce the computational complexity of whole image reconstruction. In active vision, only interest portions of the visual field are analyzed on successive frames of the image. As to gesture-based interaction, it is sufficient for gesture interpretation if spatial positions of hand, elbow and shoulder joint are identified. The conception of active vision is adopted to limit the 3D measurement of human gesture in which multiple attentive regions are created to measure the positions of hand, elbow and shoulder joint by stereo vision.

4.1 Acquisition of Attentive Regions

Human arm consists of hand, front arm and upper arm, they are connected by the joints of elbow and wrist. To determine the spatial positions of hand, elbow and shoulder joint, three attentive regions are spawned and attached on these joints in the image. A search algorithm is proposed to determine the position of hand, elbow and shoulder joint with the following procedure.

1. Acquire the segmentation image with the object of upper body.
2. Detect the position of shoulder joint by top-bottom and bottom-top search based on the width of upper body in the image.
3. Create the ray starting at the center of shoulder joint in the image, rotate the ray to search the direction that the maximum part of ray is fallen into the segmentation region. Locate the position of elbow along with the direction of ray.
4. Draw the circle at the position of elbow, detect the arc segments around the circumference. Enlarge the circle to track the arc segment of front arm until no arc segment of front arm is left, locate the position of hand at the center of last arc segment.

Figure 6 shows the principle of searching the positions of shoulder joint, elbow and hand.

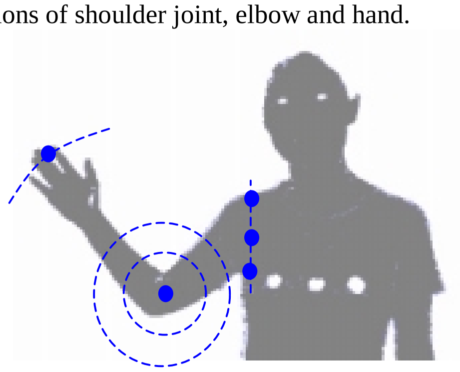

Figure 6 Location of shoulder joint Joints

4.2 Estimation of Joint Angles

The estimation of joint angles is implemented in active stereo vision, left and right images are firstly converted into LOG binary images. Three attentive regions are correlated to acquire their 3D measurements between left image and right image by using XOR operator [19]. Figure 7 shows the stereo pairs of LOG images with attentive regions.

The vectors along with upper arm and front arm are determined based on 3D measurements of attentive regions, joint angles can be easily estimated by computing the included angles among vectors.

4.3 Gesture Interpretation

The gesture of human upper body is determined by the geometry appearance of arm in 3D space. Two joint angles are defined as feature parameters to interpret the spatial orientation of arm. Thresholding of joint angles are able to classify the gesture based on threshold values of joint angles. For example, if $40° < \alpha < 135°$ and $135° < \beta < 255°$, human arm has the pointing gesture. If $\alpha < 40°$ and $135° < \beta < 255°$, human gesture is in the state of relax.

5. Experimental Results

The approach of gesture-based human-service robot interaction has been integrated into the control system of JINGANG humanoid service robot, it is verified to perform the service task of moving to the object that human points to. To implement this task, vision system is required to identify the pointing gesture and search the object along with the pointing vector. The pointing vector is a three-dimensional vector located from the center of shoulder joint to the center of elbow, object searching is carried out in a cylindrical space surrounding the pointing vector until the maximum ratio of object texture is encountered.

In the experiment, service robot attempts to recognize the pointing gesture and move its hands to the bottle which human points to. Experimental results demonstrated the effectiveness of proposed approach in gesture-based human-service robot interaction.

6. Conclusions

Gesture-based human-service robot interaction allows service robot to execute service tasks more effective and safer by using human natural communication tendency, it is essential for service robot to cooperate with people in their daily life. In this paper, an approach of gesture interpretation is presented for gesture-based human-service robot interaction. A geometry model of human body is built for gesture interpretation by analyzing the anatomy of human body, gesture segmentation is implemented by RCE neural network based on the colors of human skin and outer clothing. The view of active vision is adopted to measure the spatial position of human arm in stereo vision, three attentive regions are spawned to estimate the orientation of human arm, and human gestures are identified by the angles of shoulder joint and elbow.

The task of moving-to-point is a typical service task that pointing gesture is involved to guide the service robot close to the object, it is performed by JINGANG service robot based on pointing gesture interpretation.

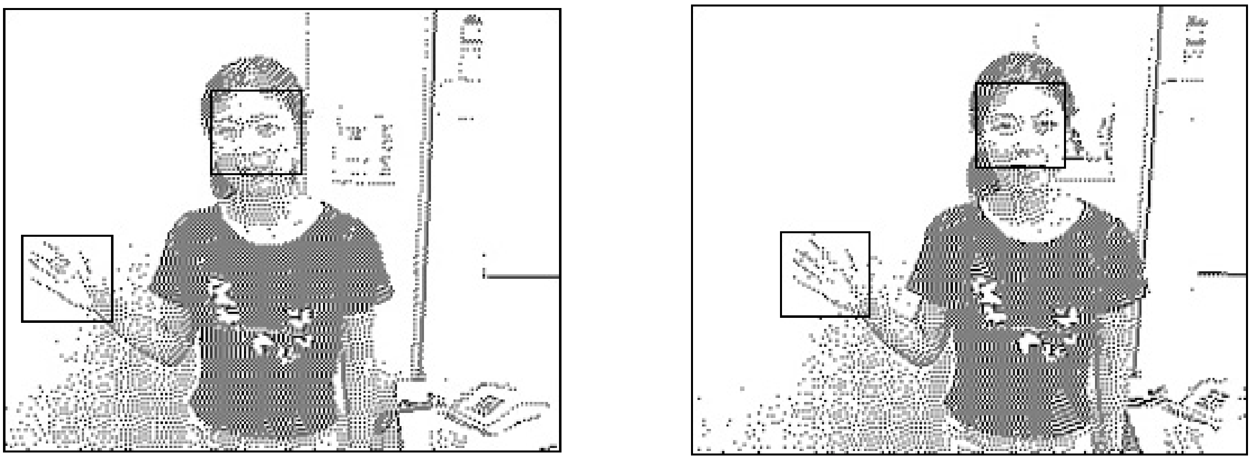

Figure 7 Stereo Pairs of LOG Image with Attentive Regions

References

[1]. Bekey G.A. Needs for Robotics in Emerging Applications: A Research Agenda. *IEEE Robotics & Automation Magazine*, 4(4):12-14, Dec. 1997.

[2]. Kawamura K., Pack R.T., Bishay M. and Iskarous M. Design Philosophy for Service Robots. *Robotics and Autonomous Systems*, 18:109-116, 1996.

[3]. Rosheim M.E. In the Footsteps of leonardo, *IEEE Robotics &Automation Magazine*, 4(2):12-14, 1997.

[4]. Brooks R.A. From Earwigs to Humans, *Robotics and Autonomous Systems*, 20:291-304, 1997.

[5]. Bischoff R., HERMES-A Humanoid Mobile Manipulator for Service Tasks, *Proc. Of 1st International Conference on Field and Service Robotics*, Canberra, Australia, 508-515, 1997.

[6]. Kaplan G., Technology 1998 Analysis & Forecast-Industrial Electronics, *IEEE Spectrum*, 73-76, 1996.

[7]. Ang W.T. and Xie M. Mobile Robotic Hand-Eye Co-ordination Platform: Design and Modeling, *Field and Service Robotics*, Springer-verlag, 308-314, 1998.

[8]. M.Ejiri, Towards Meaningful Robotics for the Future, *Proc. Of the International Workshop on Biorobotics: Human-Robot Symbiosis*, 5-6, 1995.

[9]. Myron W. Krueger, *Artificial Reality* II, Addison-Wesley, 1991.

[10]. Markus Kohler and Sven Schroter, A Survey of Video-based Gesture Recognition – Stero and Mono Systems –, *Research Report No. 693/1998*, Fachbereich Informatik, Universität Dortmund, Germany, 1998.

[11]. V. I. Pavlovic, R. Sharma and T. S. Huang, Visula Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 19(7): 667~695, 1997.

[12]. R. E. Kahn, P. N. Prokopowicz and R. J. Firby, Gesture Recognition Using the Perseus Architecture, *IEEE Conf. On Computer Vision and Pattern Recognition*, 734-741, 1996.

[13]. S. Waldherr, S. Thrun, R. Romero and D. Margaritis, Template-Based Recognition of Pose and Motion Gestures on a Mobile Robot, *Proceedings of the 15th National Conference on Artificial Intelligence*, 977-982, 1998.

[14]. D. Kortenkamp, Eric Huber and R. P. Bonasso, Recognizing and iterpreting gestures on a mobiles robot, *Proc. Of the 13th*

National Conference on Artificial and Intelligent, 915-921, 1996.
[15]. Joshua Bers, A Body Model Server for Human Motion Capture and Representation, Presence , 5(4), 381-392, 1996.
[16]. D. R. Houy, Range of Joint Motion in College Males, *Proc. of the conference of the Human Factors Soceity '83,* 1, 374-378, 1983.
[17]. D. L. Reilly, L. N. Cooper and C. Elbaum, A Neural Mode for Category Learning, *Biological Cybernetics*, 45, 35-41, 1982.
[18]. D. H. Ballard, Animate vision, *Artificial Intelligence*, 48(1), 57-86, 1991.
[19]. H. K. Nishihara, Practical Real-time Imaging Stereo Matcher, *Optical Engineering*, 23(5): 536~545, 1984.