APPLICATION TO POINTING MOVEMENT SYSTEM FOR HUMAN-ROBOT INTERFACE USING FOREGROUND SEGMENTATION WITH GAUSSIAN MIXTURE MODEL

Yihsin Ho, Yoshihiro Yamashita, Eri Sato-Imokawara, Toru Yamaguchi, Norio Tagawa
Faculty of System Design, Tokyo Metropolitan University
6-6, Asahigaoka, Hino, Tokyo, Japan

hoyihsin@tmu.ac.jp; yamashita-yoshihiro@ed.tmu.ac.jp; eri@tmu.ac.jp; yamachan@tmu.ac.jp; tagawa@tmu.ac.jp

ABSTRACT
Human-robot interface has become one of main research areas on robotic systems. The authors propose an application to Pointing movement system for human-robot interface, and focuses on developing an intuitive operation using a new manner. Our application utilized a stereo camera to capture human’s images, and utilize them as primary sources of information retrieval. We apply Foreground segmentation with Gaussian Mixture Model (GMM) for fingertip recognition to be the main image processing method to propose system. The system is mainly executed in the following steps: a stereo camera captures the user’s hand motion, the system recognizes the user’s instruction from the hand motion and the situation (as well as environmental information), and then, control a robot motion. In this paper, we describe our applied image processing method, which employs foreground segmentation with GMM for fingertip recognition, and overall concepts of the proposed application is then also defined. We present a robot demonstration to specify proposed system. Finally, the demonstration experiment is presented to show the effectiveness and research potential of our application.

KEY WORDS
Service robot, Vision, Human-robot Interface, Gaussian Mixture Model.

1. Introduction

Many robotic systems have been developed in these years. Nowadays, the concept of robotic systems is not only designed for industrial purpose but also for personal services. Since personal electronic device became popular, developers recognize the possibility of expendable personalized service systems. Through combination of concept of personalized service systems and robotic systems, the robotic systems for personal services become possible. However, in many cases, a user still has to spend long time for learning how to use the operation of a system before actual use. Thus, such a field of simple human-robot interface starts to be notice and taken seriously. Recently, intuitive operations that allow users intuitively use a system have become a popular topic in this field. Moreover, a robot in order to interact with users effectively must be able to obtain a variety of information, by detecting facial expressions, sounds, movement, etc. where human movement is one of the most commonly information for interacting the robot.

In this paper, the authors propose an application to recognize hand motion for human-robot interface, and aim to provide an intuitive operation for user. The application is to detect hand’s pointing movements, recognize the user’s instruction from hand motions, and adapt to environment information to provide personal services. Our work based on Pointing movement system [1], which is developed by Yamaguchi laboratory, to interact with a robot. This system recognizes a user’s pointing movement with following steps: a stereo camera capture user’s hand motion, the indicated point is calculated from the captured image, the system recognize the user’s instruction from indicated point and the situation (environmental information), and then, control a robot motion. However, the previous system has problem, for example: the system recognizes human’s body movement, cannot recognize subtle movement such as finger movements. Moreover, we set up the stereo camera on the robot in the previous system, so the gaze area of robot is different from user’s own. On the other hand, AR (Augmented Reality) technic displays the virtual object on the same vision. The AR system provides interactive interface that a user obtains information as if accessing real objects. However, general AR system only overlaps the information, so users cannot click or touch objects. Hence, we have been developing novel interface with the camera like eyeglasses. The camera captures user’s hands and user’s front view. This interface is suitable for robot application. In this research, we describe the implementation of the interface for robot applications, and discuss problems.

1.1 Related Work

There are researches described about capturing user’s motion for human-robot interface, such as Guanglong Du and Ping Zhang [2] used Kinect sensor, Tetsuya Kaneko et al. [3] used ZigBee sensor network. The research, which used Kinect sensor, was able to capture clear motion of limbs, but unable to capture hand’s subtle movement, no guarantee to personal privacy, and environment information ignored. The research, which used ZigBee sensor network, was able to capture user’s
precise movement and location. However, the information that capture from ZigBee sensor network is able to protect personal privacy effectively but can’t obtain environment information. Thus, in order to obtain the required information, we opted to take video by stereo camera as the main source of information retrieval for keeping environment information, and do image processing and analysis for captured user’s images.

As vision-based human motion capture and analysis are not a new research area [4], image processing methods have also been developed for a long time. Therefore, for what we ask is the method which is able to clearly capture user’s movement and adapt to Pointing movement system for human-robot interface. Specifically, the method is required to satisfy the following conditions: is able to obtain foreground and background images clearly, cost little calculation and less power consumption. There are studies proven effective methods of foreground extraction, such as Markov chain approach [5], a novel and improved method using Markov Chain theory for extracting foreground regions of human movement from a complex background scene, and Gaussian mixture model approach [6]-[8], a precise and stable method using multiple block size GMM to distinguish foreground and background regions from the captured human movement video. For example, Juhua Zhu et al. [9] present a transfer domain approach to capture foreground segmentation in real-time from a video sequences. Chris Stauffer and W.E.L Grimson [10] present a real-time tracking using Gaussian distributions as a method to modeling each pixel and evaluate to determine the background regions.

1.2 Our Approach

In this research, the authors choose GMM approach, which costs less calculation and stable for the interface. We apply the image processing with GMM for our wheelchair type robot. The proposed interface used a camera like glasses. The camera glasses captured user’s hand motion and environment information, as same as user’s vision. Thus user is able to interact with robot more intuitively.

2. The Image Processing and Recognition Method to Our Application

In this paper, the authors applied a method developed by Yoshihiro Yamashita el al.[11], which utilizes a less calculation foreground segmentation with Gaussian Mixture Method (GMM) to recognize human’s fingertip with high precision. The method uses GMM approach to distinguish foreground and background regions from the captured human movement video, and the method also mention a simple manner about calculation the indicated point from human’s finger.

The framework in GMM approach has a multiple block sizes GMM and a fine-to-coarse approach as its novel components. The multiple block sizes GMM is used to achieve different spatiotemporal performance. A small block size mode offers precise but unstable results. A large block size mode gives stable but rough result. The precise and stable performance is accomplished by combining these results. The coefficient sets of the vertical, horizontal and diagonal directions are used to determine Walsh spectrum feature parameters. Every coefficient has strong spatial correlation within the set. The fine-to-coarse is performed by introducing Walsh spectrum feature parameter. In accordance with the Walsh transfer spectrum, Walsh spectrum feature parameters are taken as the feature parameters of GMM. These feature parameters are integrated to perform the foreground segmentation. Background and foreground Gaussian distribution models are generated by modeling the background to multiple Gaussian distribution using each feature parameter of GMM. A background is determined if it matches the background Gaussian distribution model, or it is a foreground. Figure 1 shows the briefly schematic diagram of implement process of foreground segmentation with GMM.

The hand motions come from a series of hand movement. Thus, a series of images, which is a video, of the human’s hand movements are taken. These images are processed by adapted image processing method one by one. The user’s hand may be considered as the main foreground. Thus, the hand images, which only include the user’s hand, are obtained. The hand images are utilized to calculate the indicated point from a user’s finger. The recognized indicated point becomes one of the important instructions to our application. In this paper, the application recognizes a hand gesture, which is “Click motion”. Recognition results come from calculating the weight point of palm and fingertips. The application adapts the instruction and situation to control the robot motion. Click motion is recognized as following steps: images of hand motions are captured by camera, user’s hand is recognized as foreground, images of foreground is calculated to find the weight point of palm, fingertips recognized depending on the weight point of palm, and then, the Click motion is recognized by the movement from a series of hand motion images. The simple concept of calculating the indicated point from a user’s hand motion is shown in Figure 2.
3. Application to Pointing Movement System for Human-robot System

This section describes the overall concept of our application, and shows the robot demonstrations. We use two types of robots to show two demonstrations. Moreover, in order to achieve intuitive interface using AR system, our application utilizes a wearable I/O device to capture the images of hands in user’s vision. The operation manner of our application is able to capture the information that user actually seen, the precise hand motions, especially the subtle movement, also is obtained.

3.1 System Structure

Our research is based on Pointing movement system[1], thus we accords with this system to construct our application. For easily describing the system structure, we divide the previous system into three parts: input part, server part, and output part. The input part mainly includes the camera, the output part is the robot, and the server part is a computer, which is utilized as a server. The input part captures human’s images. The server part contains most of the processes of data processing and analysis, for example: image processing and analysis, motion capture and recognition, information capture, etc. Moreover, the instruction to control the robot is also recognized in here. The output part executes the instruction, which is send from the server part.

In our application, a camera glasses is utilized as input device, a robot is utilized to be the output device, and we adapt a notebook as server. The system implement process is roughly divided in to the following steps: Camera glasses capture images as primary sources of information retrieval. Server processes image by using GMM approach, the images are utilized to recognize user’s instruction and the situation of environmental. Robot moves according to the instruction from server. The system structure and system implement process as shown in Figure 3.

The robot, which is adapted in our application, owns many devices to achieve different functions. Thus it owns a robot system, which is built using RT-Middleware (RTM) [12]. RT-Middleware is software platform for building a robot system (RT system) easily. RTM is capable of combining multiple soft modules of robot functional element (RT element). RT element means the element of a robot that offers an integrated function. For example: a sensor that set up on an assistant robot is a RT element. Moreover, a control algorithm or an image processing method is also can be considered as a RT element. In simple terms, RTM combine these modularized RT elements hierarchically to build robot system easily. In this research, two adapted robots applied the RTM to build their own systems. The simple RTM system concept of two adapted robot is shown in Figure 4.
3.2 The Robot Demonstration of the Application

For specifically showing the application concept of this research, we adapt a camera glasses as shown in Figure 5 to capture the image of user’s hand motions. And we utilize a notebook, which is CF-J10 of Panasonic, to be the server for the research convenience. Moreover, we adapt a robot on the demonstration, which is a handholding robot, as shown in Figure 6. It is a wheel chair type robot.

3.2.1 The Robots for Demonstrations

The handholding robot, which is named “Buddy” and made by Vector Co., Ltd., includes a tablet, an Xtion, and a LRF. These devices are shown in of Figure 6. The set tablet is named Nexus 7, which is a production of ASUS. The camera sensor is named Xtion, which made by ASUS, it can detect human’s movement from human’s trunk and limb. The LRF adapted the production that is named URG-04LX-UG01, which is made by Hokuyo Automatic Co., LTD., it can detect existed items at surrounding area. The adapted LRF can detect the area from 0m to 5.6m, and angle measurement is 0 to 240 degrees. Xtion is able to capture user’s body movement using infrared ray. On this robot, the set Xtion is utilized to detect the user’s movement, the set LRF is use to detect surrounding area of the robot. On this robot, the set tablet is utilized to communicate to the user. Thus the tablet is able to speak and show a facial expression to the user. The facial expresses examples are shown in right side of Figure 6. On the robot, the set Xtion is applied to capture the user’s motion, the set LRF is used to detect the user’s position and obstacle on its way. Moreover, because the robot is a wheel chair type robot, the user is able to use it as a wheel chair to move. Thus, Shota Mitsumura et al.[13] has proposed a research about using this robot to assist the user who is difficult to walk. In this research, we focus on the move function of the handholding robot.
3.2.2 Demonstration Experiments

There is a demonstration experiment of the handholding robot. In the experiment, the user wears the camera glasses and is setting on the handholding robot to make an instruction using hand motion. The user instructs the handholding robot to move. We design a 3*3 trapezoidal grid for the application to decide the direction and distance, as shown is Figure 7. Moreover, because we want to provide an intuitive operation manner that user doesn’t have to learn the interface, the designed trapezoidal grid is for this application to make judgment. Thus, the trapezoidal grid is invisible to user. The application assumes human look ahead. He/she’s head is slightly down that the degree between ground and camera is about 75 degrees. The application provides 3 kinds of move distance, which are near distance about 3m, middle distance about 4m and far distance about 5m. And it provides 3 kinds of direction, which are to go forward, left side and right side. The go left side and right side mean the robot’s move direction slightly turns about 20 degrees to left or right, don’t means turn left or right. When user want to go right side and middle distance, he can just “Click” the right side and the middle space from his/her gaze area, and than, the robot will slightly turn about 20 degrees to right and move 4m. The action is executed as following steps: the user raises his hand to indicate the place in front of him by a hand motion “Click”, and put down his hand, and then, the robot move. The experiment is shown in Figure 8. Moreover, the robot is installed emergency prevention. When obstacles block on robot’s way, it will stop automatically.

The experiment shows the intuitive operation manner of proposed application, it also shows that foreground segmentation of GMM for fingertip recognition is suitable to utilize in this application. Moreover, the demonstration experiment shows the application is successful to combine a pointing movement system with a robot system.
4. Conclusion

In this paper, the authors proposed an application to Pointing movement system for human-robot interface. This paper adapted a method, which is foreground segmentation with GMM for fingertip recognition. In accordance with this method, the authors show a method of hand’s image capture from a video, and we recognize the user’s inducted point by adapting the fingertip recognition method. We proposed an improvement manner of using subtle movement capture in the pointing movement system. Furthermore, a robot demonstration is constructed. The demonstration shows the successful of combining Pointing movement system and a robot system. The combined system interface utilized the information from user’s gaze area. According to the demonstration, the authors develop an AR technic like operation manner for human-robot interface. From the demonstration experiments, we show the effectiveness and research potential of proposed application. In the future work, some topics will be continually discussed. For example: the accuracy improvement of image processing method, accuracy improvement of inducted point recognition, extendibility of the developed system and etc.

References

[1] Eri Sato, Aika Nakajima, Toru Yamaguchi, Fumio Harashima, “Humatronics(1)-Natural inter action between human and networked robot using human motion


396