Multiple People Gesture Recognition for Human-Robot Interaction

Seok-ju Hong, Nurul Arif Setiawan, and Chil-woo Lee*

Intelligent Image Media & Interface Lab, Department of Computer Engineering,
Chonnam National University, Gwangju, Korea
Tel.: 82-62-530-1803
seokju@image.chonnam.ac.kr, arif@image.chonnam.ac.kr,
leecw@chonnam.ac.kr

Abstract. In this paper, we propose gesture recognition in multiple people environment. Our system is divided into two modules: Segmentation and Recognition. In segmentation part, we extract foreground area from input image, and we decide the closest person as a recognition subject. In recognition part, firstly we extract feature point of subject's both hands using contour based method and skin based method. Extracted points are tracked using Kalman filter. We use trajectories of both hands for recognizing gesture. In this paper, we use the simple queue matching method as a recognition method. We also apply our system as an animation system. Our method can select subject effectively and recognize gesture in multiple people environment. Therefore, proposed method can be used for real world application such as home appliance and humanoid robot.

Keywords: Context Aware, Gesture Recognition, Multiple People.

1 Introduction

Recently, People prefer new input method such as eye blinks, head motions, or other gestures to traditional computer input devices such as mouse, joystick, or keyboard. Gesture recognition technology is more important than any method since it support instinctive input method. Also it is useful in multiple people environment for home appliance. Currently, there are no researches which focused on gesture recognition in multiple people situation. Most researches are focusing on gesture recognition in single person and multiple people tracking.

First we describe multiple people tracking technology. Multiple people tracking research consist of deterministic and stochastic method. In deterministic method, objects are modeled by color histogram representation, texture, appearance and objects shape such as edgelet. And then tracking is performed by matching process in hypothesized search area [1-4]. This method has a disadvantage that object's movement is fast or discontinuous. Stochastic method use probability to estimate new position of objects based on certain feature [5-7]. But this method needs a lot of computational cost so the numbers of people tracked is limited.

* Corresponding author.

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 625–633, 2007.
© Springer-Verlag Berlin Heidelberg 2007

Next we describe gesture recognition technology. Skin color based method use only skin information [8]. But it has a disadvantage that skin extraction will fail in case of complex background and illumination change. Contour based method use distance from body center point to both hand points for recognizing gesture [9]. This method is limited to recognition's number since it only uses distance information. 3D based method use 3d model of human body [10]. But it has a disadvantage such as complicated calculation cost and large database construction. Most of these works only focused on single person gesture recognition.

In this paper we deal with gesture recognition in multiple people. First of all we will define gesture and context. In segmentation part we process multiple people tracking and subject decision. In recognition part we extract feature point of body in decided subject. For extracting feature point we use two methods such as contour based method and skin based method. For recognizing gesture we use queue matching method. We also introduce animation system as an application. Finally we will show experimental result and conclusion. Our system architecture is shown in Fig. 1.

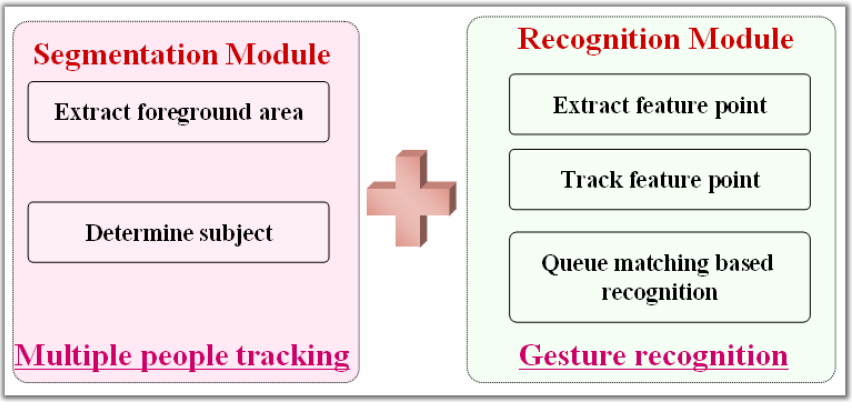

Fig. 1. System architecture (Segmentation module and Recognition module)

2 Context Aware Gesture Definition

In this section we define gesture which is used in our system. Next, we define each individual person's state from input image. Finally we describe state transition model for selecting subject.

2.1 Definition of Gesture

Mankind expresses his mind using eye blink, body movement, or sound. Specifically both hands' movement is used for expressing gesture. So gesture can be analyzed by using movement of both hands. We can not define all gestures used by people. Therefore, we define five gestures for human-robot interaction as shown in Fig. 2. Each gesture meaningfully separated into each other.

Fig. 2. Definition of Gestures (come here, stops, shake hands, heart, bye bye)

2.2 Definition of State

"Context" is consisting of many situations such as illumination, number of people, temperature, noise and so on. In this paper, we define the "context" as intention state between human and computer. We selected speed and distance as intention of behaviors. According to these factors, state will be decided as shown in Table 1. The speed decides [Walking] or [Running] and the distance of behavior is most important factor since it decides to apply gesture recognition algorithm.

Each person has an only one state per every frame. Each state change using state transition model as shown Fig. 3. We assume that there are 3~4 people in input image. If one person going closer, we decide the person as a subject. If a subject decided, we extract feature point from subject's area. In next section, we describe how to extract feature point and how to recognize gesture.

Table 1. Definition of states

State	Definition
Null	Person is not exist.
Object	Person is detected.
Walking	Person is detected, and he/she is walking.
Running	Person is detected, and he/she is running.
Recognition Enabled	Person is detected, and he/she is closer or doing specific gesture.

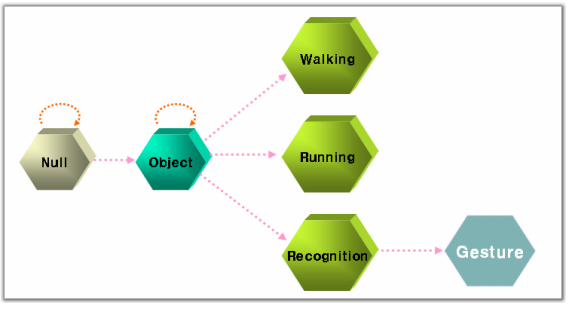

Fig. 3. State transition model of our system

3 Feature Extraction Method

In this paper, we extract area of both hands and head. Segmentation process use Gaussian mixture model in improved HLS space [11]. We use two methods for extracting feature area. First method is contour based method. Second method is skin based method. In this section, we describe these methods and tracking method.

3.1 Contour Based Method

In segmentation process, we extract subject's silhouette from input image. We must eliminate noise since silhouette image has a many noise. To remove this noise we apply dilation operation as shown equation 1. Contour line data is easily extracted from binary image data. We use OpenCV library for extracting contour. It retrieves contours from the binary image and returns the number of retrieved contours. We can get contour line to connect retrieved contour points. Contours can be also used for shape analysis and object recognition.

$$A \oplus B = \{ z \mid [(\hat{B})_z \cap A] \subseteq A \} \tag{1}$$

Fig. 4. Contour based method (input image, segmentation result, feature point result, wrong extracting feature point)

After extracting contour, we extract feature point for using contour based method. First we define three points of body (Left Hand-LH, Right Hand-RH, and Head Point-HP). [LH] point is the lowest X coordinate of contour result. [RH] point is the highest

X coordinate of contour result. [HP] point is the lowest Y between [LH] and [RH]. Extracted points will use for recognizing gesture.

This method has an advantage that calculation cost is simple. But these extract wrong points since both hands are occluded in body area as shown in Fig. 4. To solve this problem we must estimate points when position of both hands is change quickly.

3.2 Skin Based Method

Skin is an important factor for extracting both hands and head. There are many methods how to extract skin from image. In this paper we use to extract skin from YCBCR image. First of all, we apply mask in segmentation silhouette image and then we can get only subject area. And then we convert masked RGB image into YCbCr image. If we apply defined threshold in YCbCr image, we can get skin result image as shown Fig. 5.

Fig. 5. Skin based method (input image, segmentation result, masked segmentation image, extracted skin result image)

For recognizing gesture we must decide both hands position from skin result image. Both hands position can get x-y coordinate from x-y projection. Intersection of x projection and y projection is position of both hands and head. Our result can show in Fig. 6.

Skin based method has lower calculation cost than contour based method. Also this method can detect both hands points when both hand s occluded. But this method arise problem when illumination change. Also this must apply another skin threshold for different human race.

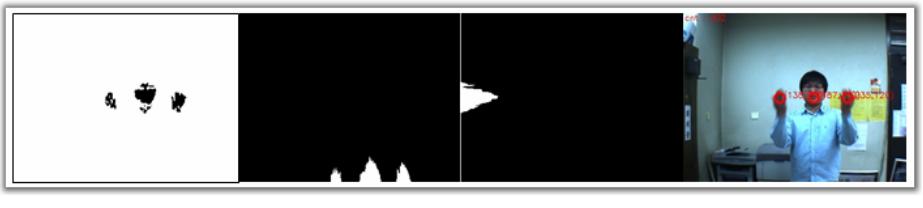

Fig. 6. Extracted skin result image, x projection image. y projection image, feature point

3.3 Feature Tracking Using Kalman Filter

In this paper, we use a Kalman filter for tracking both hands. The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive)

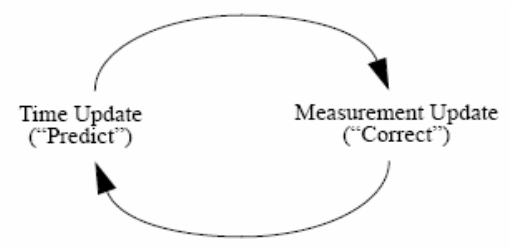

Fig. 7. Kalman filter algorithm architect

solution of the least-squares method. The filter is very powerful in several aspects: it supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown.

The Kalman filter estimates a process by using a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of (noisy) measurements. As such, the equations for the Kalman filter fall into two groups: time update equations and measurement update equations. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. Indeed the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems as shown below in Fig. 7.

4 Gesture Recognition Using Queue Matching

The gesture contains user's intentions in motions of whole body. Especially, trajectories of hands include more intentions. So, we adopt the different recognition

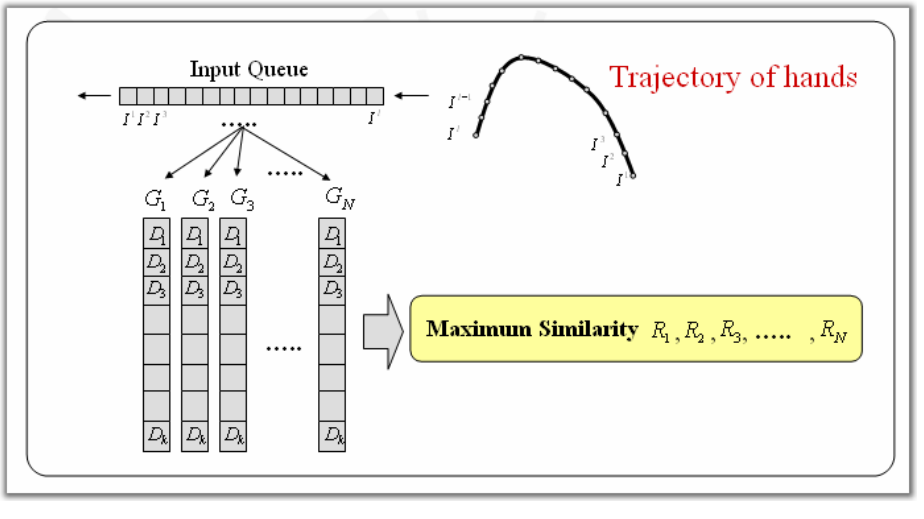

Fig. 8. Queue matching method for recognizing gesture

method which uses trajectories of hands as features. Many researchers have tried to develop the matching algorithm for the trajectories in a number of ways. Generally, the methods are used for recognition of handwritten character. But, it is not effective to apply into the gesture recognition, because it is difficult to decide the start and end point of meaningful gestures. Therefore, many researchers are continuing to study about the problems, Gesture Spotting [9].

In this paper, we propose the simple queue matching method instead of gesture spotting algorithm if the gestures are not complicated. And this method has the advantage in fast to process and easy to implement.

The basic concept of this algorithm is as follows. Assume that the model set M has N models. Also, direction vectors represent the trajectories of hands, and these vectors are stored continuously in each gesture models.

We can get directional vectors from each frame. And, input queue with the length I is a set of these vectors. If the meaningful gesture of subject is in the input queue, it can be assumed that this queue includes the subject's intention. And then, input queue is compared with each model gesture. Finally, we can decide the gesture, as a recognition result. In next section we introduce our system as an application.

5 Application: Animation System

In this paper we use our system as an animation generation system. From input image we construct 3D body model in virtual space. 3D body model has a similar appearance with subject. Also this model has a similar action with subject's action.

To construct animation system, we use feature point from gesture recognition system. These points used for estimating human body point. Extracted feature points have many noises from general environment. We use NURB algorithm for eliminating noise. And we estimate each body joint position using Inverse Kinematics. To estimate correctly, we use information such as human anatomy, previous frame information and collision process. Finally, we estimate body point using extracted feature point and end-effector.

To represent 3D model, first we construct 3D virtual space in animation system. Gesture recognition system send to animation system feature point information. We can get animation system similar doing input gesture.

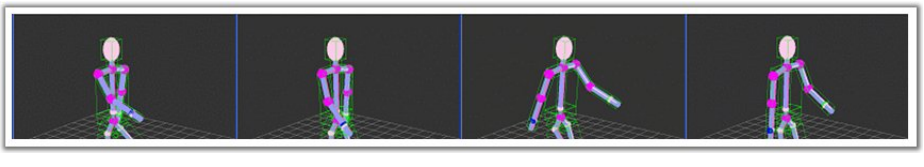

Fig. 9. Implemented animation system

6 Future Work

The experiment was taken on 2 PCs with 3.0 GHz Intel Pentium 4 CPU and 512MB RAM. We used Bumblebee of Point Grey for extracting stereo information. The

system is written in Visual C++ 6.0 based on OpenCV 1.0. Fig. 6 shows results of extracted feature points and gesture recognition result.

Contour based method has a problem when both hands are occluded in body area. For example, both hands position go wrong when [heart] and [bye bye]. Skin based method extract good position in every gesture. It is shown robust result when both hands are occluded in body area. But skin is failed when illumination change.

We have a problem since we use only 2 dimensional data for recognizing gesture. For example, we can not recognize both hands upward and both hands upward in round fashion. This can recognize gestures if we use 3 dimensional data instead of 2 dimensional data. And our system can not make trajectory information when subject doing [shake hands gesture] and subject doing [bye bye gesture]. To solve this problem, we must use time information and movement information of specific area. If we use a convex hull algorithm for extracting feature point, we can have a simple calculation cost and accurate feature points.

Fig. 10. Contour based gesture recognition result(come here, stop, shake hands, heart, bye bye)

Fig. 11. Skin based gesture recognition result(come here, stop, shake hands, heart, bye bye)

Also we have a problem when subject is changed. Subjects have a little different trajectory information of gesture. To solve this problem, we assign a personal ID. Our system recognizes a personal ID, and it uses a model gesture of ID as a model gesture.

In this paper, we proposed gesture recognition in multiple people environment. Our system is divided into two modules – segmentation module and gesture recognition module. Also our system can change subject if subject entered. And then our system tracked feature points using Kalman filter. Finally, our system can recognize gesture using simple queue matching.

In this paper, we propose animation system using implemented gesture system. This system can make 3D information of human. We can get automated animation in future.

Our method can use general interface of robot. If it solve previous problem, intelligent robot can communicate with mankind naturally.

Acknowledgments. This research has been supported in part by MIC & IITA through IT Leading R&D Support Project and Culture Technology Research Institute through MCT, Chonnam National University, Korea.

References

1. Zhao, T., Nevatia, R.: Tracking Multiple Humans in Crowded Environment. In: Proceedings of CVPR 2004, pp. 1063–6919 (2004)
2. Wu, B., Nevatia, R.: Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. In: Proceedings of ICCV 2005, vol. 1, pp. 90–97 (2005)
3. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: Real-Time Surveillance of People and Their Activities. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 809–830 (2000)
4. Siebel, N.T, Maybank, S.: Fusion of Multiple Tracking Algorithms for Robust People Tracking. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 373–387. Springer, Heidelberg (2002)
5. Franc, J.B., Fleuret, o., Fua, P.: Robust People Tracking with Global Trajectory Optimization. In: Proceedings of CVPR 2006, vol 1, pp. 744–750 (2006)
6. Nguyen, H.T., Ji, Q., Smeulders, A.W.M.: Robust multi-target tracking using spatio-temporal context. In: Proceedings of CVPR 2006, vol. 1, pp. 578–585 (2006)
7. Han, J., Award, G.M., Sutherland, A., Wu, H.: Automatic Skin Segmentation for Gesture Recognition Combining Region and Support Vector Machine Active Learning, In: Proceedings of FGR 2006, pp. 237–242 (2006)
8. Li, H., Greenspan, M.: Multi-scale Gesture Recognition from Time-Varying Contours. In: Proceedings of ICCV 2005, vol 1, pp. 236–24 (2005)
9. Lee, S-W.: Automatic Gesture Recognition for Intelligent Human-Robot Interaction. In: Proceedings of FGR 2006, pp. 645–650 (2006)
10. Setiawan, N.A., Hong, S-j., Lee, C-w.: Gaussian Mixture Model in Improved HLS Color Space for Human Silhouette Extraction. In: Proceedings of ICAT 2006, pp. 732–741 (2006)
11. http://www.sourceforge.net/projects/opencvlibrary