울산대학교 공학연구논문집 제30권 2호 pp. 359~366, 울산대학교.
Journal of Engineering Research Vol. 30, No. 2, pp. 359~366, 1999
University of Ulsan.

Vision-Based Human Interface System for Hand Manipulative Gestures

Kang-Hyun Jo
School of Electrical Engineering and Automation,
E-mail : jkh@uou.ulsan.ac.kr

Abstract

This paper presents a system recognizing manipulative gestures like grasping, moving, holding an object(s) with both hands, and extending or shortening of the object(s) in the virtual world using contextual information. Contextual information is represented by a state transition diagram, each state of which indicates possible gestures at the next moment. Image features obtained from extracted hand regions are used to judge state transition. When we use a gesture recognition system, we sometimes move our hands unintentionally. To solve this problem, the proposed system can recognize collaborative gestures with both hands. They are expressed in a single state so that the complexity in combination of gestures of each hand can be avoided. We have realized an experimental human interface system. Operational experiments show promising results.

1 Introduction

Vision-based interfaces with which we can give orders to computers by hand gestures have attracted much interests [1,2,3,4]. However, most conventional systems can deal only with pointing gestures [5,6,7]. We can point at an object in a 3-D virtual world and move by hand gestures with these systems. However, if we want to handle object parts more freely in the virtual world to design an object by combining these parts, we need various other manipulative operations, such as grasping, moving, holding an object with both hands, extending, shortening, and putting down on the desk. An interface system should recognize manipulative gestures corresponding to these operations. However, some of these gestures are similar if we look at each

separately, thus this cause difficulty in recognition only with the appearance. We propose a use of contextual information to solve this problem. We use the virtual reality system for a certain purpose. Thus, we can reduce the number of possible next operations following a particular operation to a small number. We do not take an operation that does not lead the status closer to the goal. We represent this knowledge by a state transition diagram.

Human manipulative gestures are often conducted by the two hands. Thus, vision-based interfaces should have the capability of recognizing collaborative gestures with both hands. In this system, each collaborative gestures is expressed in a single state in the state transition diagram so that the complexity in combination of gestures of each hand can be avoided.

Conventional human interface systems consider only meaningful gestures. However, we sometimes move our hands unintentionally. Human interface systems must discern such human's unintentional actions from intentional manipulative gestures and respond only to the latter. Our human interface has a rest state in the state transition diagram. If the system considers that the current human action might be unintentional because the extracted features are out of expected ranges, the system takes a rest in the rest state until it can make a more certain decision.

This paper presents our context-based manipulative hand gesture recognition method. It also describes an experimental human interface system using the proposed method.

2 Context-based approach to recognize human gesture

In this paper, we present a human interface system that enables a human user to manipulate objects in the virtual 3D world. In the virtual space, he may point at an object, grasp it, bring it into the work space, and may change its size or put it down on another object. These actions will not happen independently. We can choose possible actions(hand gestures) following each action. We represent these relationships by a state transition diagram. Since this diagram limits the number of possible recognition classes, the system can recognize the next gesture using simple image features.

Those state inferences are based on the visual observations. The observations are described in feature vectors which simplify the information of hand position and shape. According to the conditions of state inference, the current state can be determined. As shown in table 1, the states are inferred by the simple observations. For example, when the previous state is grasping, transferring to moving is only checked with area and position of hand from the feature vector. Other elements of feature vector are regarded useless for this state transition.

2.1 Feature extraction

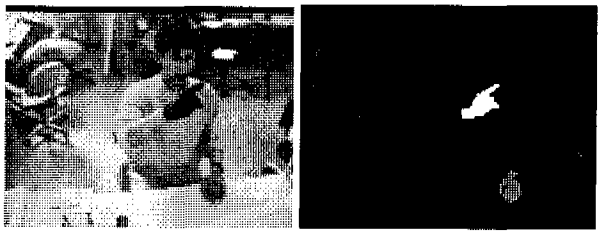
In order to obtain hand features, hand regions need to be extracted. Each hand region can be extracted by thresholding the hue information of color images. In the current implementation, we have to wear a green glove and a red one on each hand to realize fast reliable feature extraction. Fig.1 shows an example of hand region extraction.

The feature vector, $O(t)$, is calculated from this segmented hand regions. The hand feature vector consists of the following elements:

1. area : the region,
2. cx : x coordinate of the centroid of the region,
3. cy : y coordinate of the centroid,
4. fmax : the finger tip length from the centroid, which is the maximum of $f(\theta)$ indicating the distance between the centroid and the hand region contour in direction θ,
5. fx : x coordinate of fathest point from the centroid,
6. fy : y coordinate of fathest point,
7. fmean : the mean of $f(\theta)$ for all θ.

We use two cameras placed upper front of the user and above. The feature vectors are calculated for each hand in each image.

Except certain pointing gestures, we cannot recognize gestures from a frame of an instance. Thus, the system observes gestures for an appropriate length of interval determined for each state.



(a) Original image. (b) Extracted hand regions

Figure 1 Example of hand region extraction.

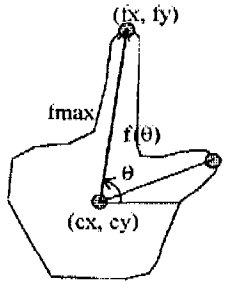


Figure 2 Computed features.

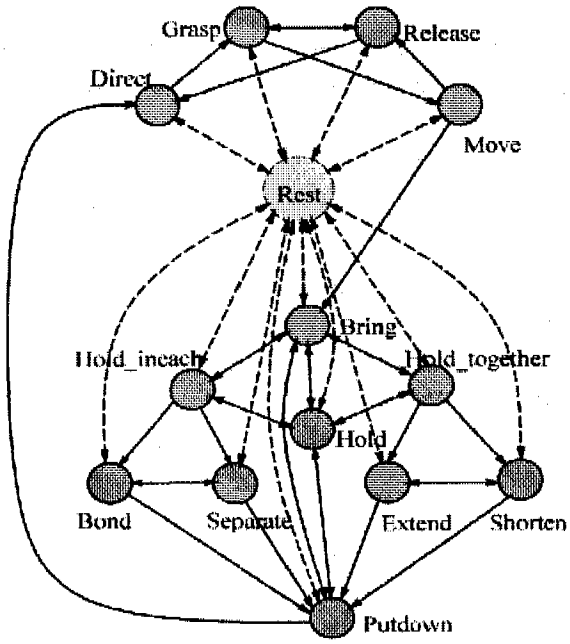


Figure 3 State transition diagram(self loops are omitted).

2.2 Gesture recognition

Fig. 3 shows the state transition diagram used in the system. Table 1 describes the conditions for state transition. Except pointing gestures such as Direct and Move, the system uses feature vectors for several frames to recognize gestures. To cope with the variation of motion speed, the system examines in most cases whether particular feature values are increasing or decreasing as shown in Fig. 4 rather than checks their exact values.

In Fig. 3, transition from and to Rest is depicted by dotted lines. If a new observation does not match any possible state transition conditions from the current state, the system changes it to the Rest state. While in the Rest state, if the system observes a feature vector that satisfies one of the possible transition conditions from the previous state, the state is changed according to the condition.

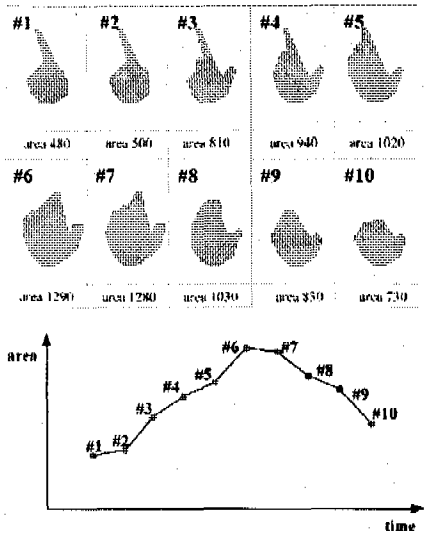


Figure 4 Example of recognition of Grasp.

2.3 Collaborative gestures with both hands

Human manipulative gestures are often conducted by the two hands. Thus, vision-based interfaces should have the capability of recognizing collaborative gestures with both hands. In this system, we assign a state for each manipulative gesture with both hands. This solves the combinatorial problem that arises when a state is given for each hand and manipulative gestures are represented by the combination of both hand states. As shown in Fig. 3, Hold_together is assigned to the manipulative gesture for holding an object with both hands. If this has been recognized, the object will be extended, shortened or released after then.

3 System configuration and operational experiments

We have developed an experimental human interface system using the proposed recognition method. Fig. 5 shows the system configuration. The current system uses two Unix machines(Sun's SS5 and SS20), each for feature extraction and graphics rendering, respectively. The latter displays two views for the user as shown in Fig. 6. One is a usual perspective view of the virtual world. The other is a top view of the virtual world to help the user to understand positional relations of objects.

We consider the following scenario for operational experiments. There is a palette space(PS) in the virtual world where object components are floating. We Direct(point at) a desired object in PS and Grasp it by hand gestures. Then, we Move it and Bring

it to the work space(WS), which is a space on the desk in the virtual world. We hold the object by both hands(Hold_together) and Extend it or Shorten it to change the object size to desired one. Then, we bring another object component from PS and put it down on the first object(Putdown).

We carried out operational experiments according to the above scenario. We were able to build such an object as shown in Fig. 6 by hand gestures even if we added any unintentional movements of the hands during operation.

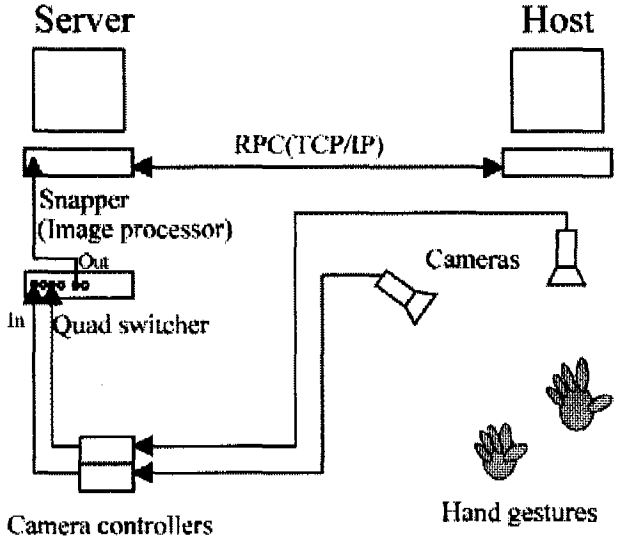


Figure 5 System configuration.

4 Conclusion

Hand gestures have a large variety in expressing human subject's intention. We assumed such gestures as those the human subject shows his intention when he manipulates objects in the virtual world. We define the gestures as states so that they show the context of behaviors meaning. Thus, the manipulative gestures were recognized on the basis of context. An experimental system shows the explicit manipulative hand gesture can be recognized and interacts with the human subject. The system runs in 2 frames per second thus needs to perform faster for the comfortable interaction. Future research remains with automatic extraction of human skin region and identification of each region of hand. Furthermore, hand and face recognition should be incorporated for the complicated and intelligent human interface system.

References

[1] T.S. Huang and V.I. Pavlovi ć, ''Hand Gesture Modeling, Analysis, and Synthesis'', Pooc. Workshop the Second Int'l Conf. On Automoatic Face and Gesture Recognition, pp.73-79, 1995.

[2] R. Kjeldsen and J. Kender, ''Visual Hand Gesture Recognition for Window System Control'', Pooc. Workshop the Second Int'l Conf. On Automoatic Face and Gesture Recognition, pp.184-188, 1995.

[3] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, ''The ALIVE System: Wireless, Full-body Interaction with Autonomous Agents'', MIT Media Lab. Perceptual Computing TR No. 257, 1995.

[4] Francis K.H. Quek, ''Unencmbered Gestural Intraction'', IEEE Multimedia, Vol.4, No 3, pp.36-47, Winter 1996.

[5] M. Fukumoto, K. Mase, and Y.Suenaga, "Real-time detection of pointing actions for a glove-free interface", Proc. IAPR Workshop on Machine Vision Applications '92, pp. 473-476, 1992.

[6] R. Cipolla, P.A. Hadfield, and N.J. Hollinghurst, "Uncalibrated stereo vision with pointing for a man-machine interface", Proc. IAPR Workshop on Machine Vision Applications '94, pp.163-166, 1994.

[7] K.-H. Jo, K. Hayashi, Y. Kuno and Y. Shirai, ''Vision-Based Human Interface System with World-Fixed and Human-Centered Frames Using Multiple View Invariance'', Trans. IEICE Information and Systems, Vol.E79-D, No.6, pp.219-228, 1996.

[8] K.-H. Jo, "Hand Gesture Recognition using Human-Centered Frame and Task Knowlege for Smooth Human-Computer Interaction", Ph.D Thesis, 1997.

[9] K.-H. Jo, "Hand Manipulative Gestures Recognition System Using Task Knowledge", Proc. ICEIC'98, pp.(II)5-8, 1998.

Index	
ⓐ	Hand is located in PS.
ⓑ	Finger tip is erected.
ⓒ	Area is getting smaller for the last 2 frames.
ⓓ	Area is getting larger for 3 frames consecutively.
ⓔ	Area was getting larger for 3 times in the last 8 frames right before the last 2 frames.
ⓕ	Area is not abruptly changed.
ⓖ	Both hands are located in WS.
ⓗ	Hand is still.
ⓘ	Distance between both hands is getting larger.
ⓙ	Hand is moving.
ⓚ	Distance between both hands is getting smaller for 3 frames consecutively.
ⓛ	Other hand is in the state of 'Grasp'.
ⓜ	Both hands are getting larger for 3 frames consecutively.
ⓝ	Both hand areas do not change abruptly.

Table 1. Conditions for state transition.

P S: Palette space where object components are floated.

WS: Work space where target objects are manipulated.

The finger tip is erected : fmax > fmean + threshold.

Area is getting larger : area > prev_area + thresh_area, where prev_area is the area in the previous frame and the thresh_area is a threshold with respect to area.

State transition	Conditions
Start -> Direct	ⓐ , ⓑ
Direct -> Direct	ⓐ , ⓑ
Direct -> Grasp	ⓐ , ⓒ , ⓔ , ⓗ
Grasp -> Release	ⓐ , ⓓ
Grasp -> Move	ⓐ , ⓕ , ⓙ
Move -> Bring	ⓐ , ⓕ , ⓙ
Bring -> Hold_together	ⓕ , ⓖ , ⓚ , ⓛ
Bring -> Putdown	ⓖ , ⓜ
Hold_together->Extend	ⓖ , ⓘ , ⓝ

Table 2. Using indices of Table 1, each transition can understanding the behavior.

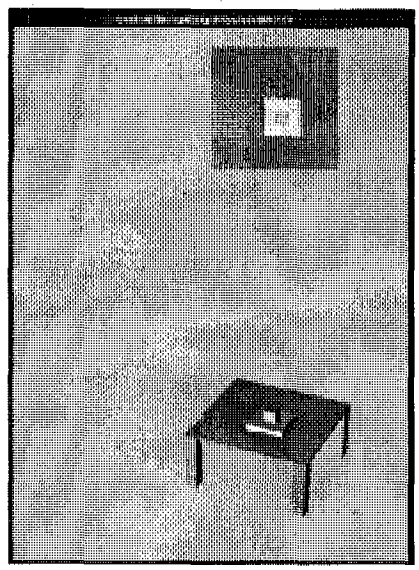


Figure 6 Display Example: a small object is located on the center of the bigger plate.