4L-7

オブジェクトとの関係による拘束を用いた関節角度の系列に基づく身体動作の解釈

ジュアンダロクマン　　金子正秀

電気通信大学　大学院電気通信学研究科

1. Introduction

Human body consists of many parts, which are connected to each other by joints (on-rigid articulated objects). It can yield many different poses from a simple movement/gesture to complex and complicated acrobatic motion in daily lives scene.

Human also can perform one activity or several concurrent activities at any given time with or without interacting with objects in the environment, e.g., walking while making phone call, walking while waving hand. One activity can be performed by single part of the body or as coordination of whole body part (whole body activities).

This paper focuses on 2 major parts, (1) interpretation of each part of human body by hierarchical scheme, and (2) considering the relationship of human with the object that he is being interacted with.

2. Hierarchical analysis of body parts

The basic ideas for analyzing each part of the body are (1) human activity employs at least one of the body parts, (2) human can carry out many activities at the same time, and (3) human can clearly understand only that can be perceived.

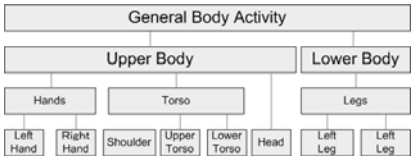


Fig.1: Hierarchical Interpretation.

As shown in Fig. 1, we analyze each body part separately at the lowest level, and combine them at higher levels, thus make multilevel interpretation. Each level of interpretation will give a recognition output. As the final output (according to priority) are whole body activity, upper/lower body activity, pair of body part activities, and finally part of body activities.

Table 1: Example of recognition result by hierarchical scheme

Test Data	**Body Part Interpretation**		
	Hand	Torso	Leg
Raising Hand	raising		
Bowing		bending	
Pick-up	forward	bending	
Walking	swing		swing/walk

Interpretation of Activities of Human Body Parts Based on Sequence of Joint Angles under Considering Constraint Relationship with Objects
Juanda Lokman, Masahide Kaneko
The University of Electro-Communications

Table 1 shows the example of recognition result by using hierarchical scheme. Single activity like rising hand and bowing just involve one part of the body part, pick-up is recognized as coordination between hand and torso, while walking could be recognized as one coordination activity that involved hand (swinging) and leg (swinging), or as major activity of leg, thus it can recognize simultaneous activity while human is walking.

Here, we will concentrate our work on motion interpretation by considering the relation with the object and we will not use any object recognition and tracking technique to recognize and track the object in the scene, thus we implement the objects in 3D virtual world. Using 3D object in virtual world, the system already know the position of the objects. Thus we only need to track the pose of the human. We construct the 3D position of human head and hands, using multiple stereo cameras, then combine into the 3D virtual world to interact/manipulate the objects.

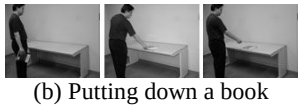
3. Objects as context

Human activities can be understood from the body pose or sequence of body poses, but in many cases while human is performing an activity, it always involves some objects in the environment. And some of the activities could have the same body pose or sequences of body poses thus result in the same perception (see Fig. 2(a) and 2(b)) if without concerning about the object that being interacted with.

Characteristic of the object also affect the activities of the human, for example one way door is opened from inside the room and outside the room result in different sequence of poses (Fig. 2(c) and 2(d)).

While human manipulating the object in the scene (e.g., picking up and putting down on the different places), the position of the object is important as well for being tracked. Thus not only the relationship between human and object is important, but relationships among objects themselves are important as well.

(a) Picking up a book



(b) Putting down a book

(c) Opening (pulling) a door

(d) Opening (pushing) a door

Fig.2: Several Motion Events

Thus, combining object information and the pose of human body can give the solid interpretation of the scene [2], instead of only rely on the pose or sequence of poses of the body.

3.1. Objects description

Objects in virtual world can be tracked while the human manipulates the object. The object that is being manipulated and its position is tracked whole time by the system.

All objects are described by the concept of class in programming language (object oriented programming) with hierarchical scheme. Object description contains the context information about the object (related activity and the position of the object not only in 3D space but the relative position of the object to other objects, e.g., glass on the table or floor). Every object is derived from the main class that self contains about the information of general properties for every object in the world. For example, basically most of the object can be moved (pick up or put down), depend on the size and weight.

4. Hands and head position tracking

Location and size of the objects in virtual world are built based on the objects in real world. The environment for experiment is done in real world and virtual 3D world. The human pose and motion is captured, and then combined into virtual 3D world, where the 3D object is made and located at the same position with the real world. Human are trying to interact with the object in the real world at the same time manipulating the object in 3D world as well.

4.1. Multiple stereo cameras

We use 3 stereo-cameras, each of which is located at left, right and front of the table respectively. Field-of-view and line-of-sight restriction can be overcome using multiple cameras or stereo-cameras. Using this configuration, most of the pose and position of the hands and head can be calculated using combination of 2 stereo-cameras.

4.2. Skin-color detection

We use skin-color detection to detect the area of the hands and head, and then 3D position of the hands and head is captured and calculated by stereo-cameras. Skin-color detection is modeled by GMM in HSV space [3].

Fig. 4 shows the result of tracking the position of hands and head using skin-color technique and multi-stereo-camera view. From input color, it detects the position of the hands and head, and stereo-cameras will calculate the position of them in 3D space. Finally it combines each correspondence hands and head from each stereo-camera. Fig. 5 shows the interaction of the left hand with the glass in real world and the tracked left hand with the object in the virtual world.

5. Conclusion

Hierarchical scheme is useful for interpretation of complex and simultaneous human activities by analyzing each part of the body part separately. Considering the relationship between human and object can eliminate the ambiguity among the activities that have the same pose or sequence of poses, thus give more reliable interpretation of activities.

Knowing about what is happening in the scene especially what human is doing, what or to whom the human interact with, position of the human(s) and object(s) are important for giving support to the human.

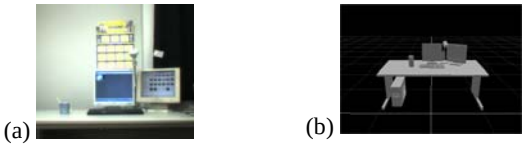


Fig. 3: (a) Object in real world and (b) in virtual world

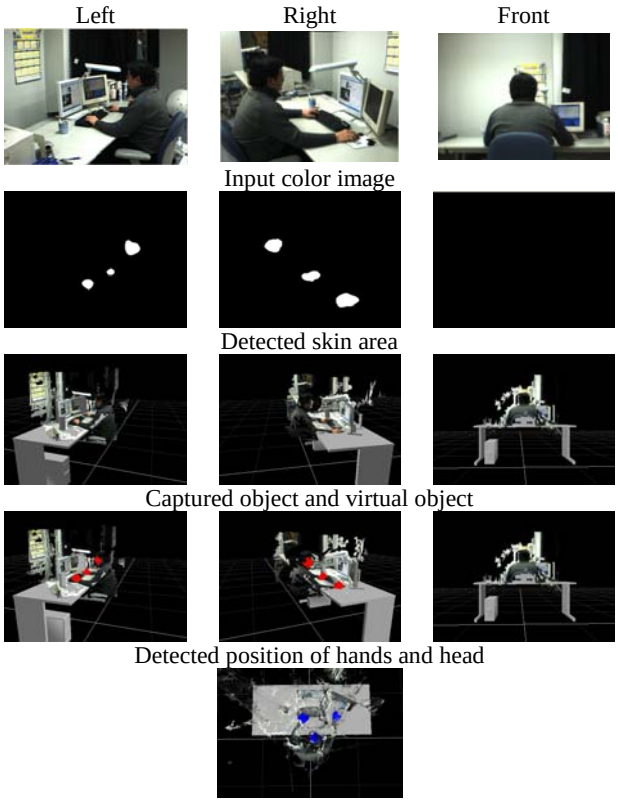


Fig. 4: Example of captured position of hands and head.

Fig. 5: Interaction with the glass in real and virtual world.

References

[1] Juanda Lokman and Masahide Kaneko: “Hierarchical Interpretation of Composite Human Motion Using Constraints on Angular Pose of Each Body Part,” RO-MAN 2004, pp. 335-340 (Sept. 2004)

[2] Douglas Ayers and Mubarak Shah: “Monitoring Human Behavior from Video Taken in an Office Environment,” Image and Vision Computing, vol. 19, no. 12, pp. 833-846 (Oct. 2001)

[3] Quan Huynh-Thu, Mitsuhiko Meguro, and Masahide Kaneko: “Skin-color extraction in images with complex background and varying illumination,” Proc. Of IEEE WACV2002, pp. 280-285 (Dec. 2002)-