Micro-pose for Gesture Recognition with Bodily-Pose Decomposition

Jae-wan Park, Su-yong Park, and Chil-woo Lee

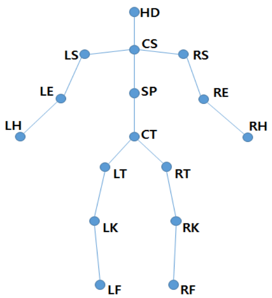
Dept. of Electronics and Computer Engineering, Chonnam National University, Gwang-ju, Korea
{cyanlip,qkrtndyd12}@naver.com, leecw@jnu.ac.kr

Abstract. Because a pose is implemented with static configuration of bodily joints expressing an abstract meaning of human intention, a gesture can be thought as a sequential combination of specific poses of human body. Therefore we can understand the meaning of a gesture if we can analyze the arrangement of the poses in spatiotemporal space. In this paper we propose a novel gesture recognition algorithm in which a pose is decomposed into micro-poses that are constructed with a small number of bodily joints. A micro-pose is a smallest unit of joints which can include a physical mean of body. To obtain the location of body joint MS Kinect is used and the information extracted from the micro-poses is finally applied to code matrix as the code book of the Bag-of-Words to understand the meaning of the gesture.

Keywords: Micro-pose, pose estimation.

1 Introduction

We can think a human gesture as a typical combination of bodily poses in temporal space and also a pose is implemented with static configuration of bodily joints expressing an abstract meaning of human intention. That means that we can understand the meaning of a gesture if we can analyze the spatiotemporal configuration of the joints of human body. In this paper we propose an algorithm which adopts the novel approach of "divide and conquer" for pose decomposition in gesture recognition.



HD	Head
CS	Center Shoulder
SP	Spine
CT	Center Thigh
LS	Left Shoulder
LE	Left Elbow
LH	Left Hand
RS	Right Shoulder
RE	Right Elbow
RH	Right Hand
LT	Left Thigh
LK	Left Knee
LF	Left Foot
RT	Right Thigh
RK	Right Knee
RF	Right Foot

Fig. 1. Human body structure with sixteen principal joints

C. Stephanidis (Ed.): HCII 2014 Posters, Part I, CCIS 434, pp. 451–456, 2014.

As we know, human body has a very complicated and articulated 3D structure of bodily joints as shown in Fig.1. So, we suggest an approach using decomposition of large set of whole bodily joint into general case of a small set of some specific joints. We call that configuration of small joint set micro-pose. If we utilize this micro-pose for pose recognition the whole process of the recognition becomes straightforward because we can make use of simple combination of micro-pose instead of complex matching process in huge search space.

In most case, pose recognition method usually can be divided into two approaches. The first approach uses appearance information of human body such as silhouette and edge. We call this approach "appearance base method" and the implementation of the method is very simple since we can obtain sufficient information if we find distinctive patterns from the appearance for a specific gesture. However, it sometimes cannot produce robust result because we cannot assume general pattern for the specific pose since the appearance is changed according to view direction and different person.

In the second approach, we call it "model based approach", the location of joints of body structure is used for recognition. In this case, we may face to solve the problem of estimating accurate locations of the joints which need much computational cost, so it was very difficult to implement algorithms in real-time. Consequently for the practical application, we used to attach some markers on the dominant joints of body.

And body part based method can be used Cardboard People [1], Pictorial Structure [2] and Poselets [3] for body representation. These recognizing method can be represented as rigid body set of the constellation. Namely, they connect the close body joints with straight lines. And they can use the connected body-joints as a tree structure. This body part based models should need two principal elements. The first one is that each body part has been combined to the space of real image, and the second is that connection of body joints has to be arranged with hierarchical system.

In this paper, we analyze features of the poses consisting of gestures according to the body part based method, and we recognize the gestures with code matrix as the code book of the Bag-of-Words.[4].

2 Micro-pose for Gesture Recognition

From the topology of the fig.1 we firstly derive the micro-poses for limbs; MP1 and MP2 are assigned to the joints connecting left and right shoulder to each hand, and MP3 and MP4 are also assigned to the joints connecting left and right thigh to each foot respectively. These MPs are used for display the deformation for limbs. In this paper, in order to configure micro-pose group, we use sixteen principal joints.

For the rotation of arms and legs around bodily axis we consider another four Micro-Poses as shown in fig.2 (e, f, g and h). As the same with the previous MPs, they are also joint groups which connect principal joint of limbs to the center of a body. These four MPs are used for dominant rotation of arms and legs.

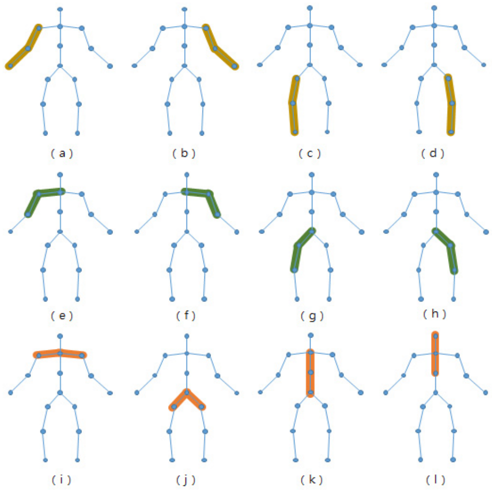


Fig. 2. The joint group used for Micro-poses

Practical shape of the MPs is decided by the angle between two joints around the center joint and the angle can be used as feature vector for pose recognition. For instance, fig.3 shows the three feature vectors projected onto X-Y, Y-Z, and Z-X plane respectively.

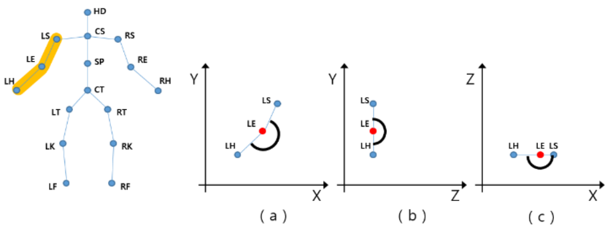


Fig. 3. Feature Vectors for Micro-pose Groups ((a) definition of the angle in X-Y plane (b) definition of the angle in Y-Z plane (c) definition of the angle in Z-X plane)

3 Gesture Recognition

In this paper, we use MS KINECT[5] to obtain the 3D information of bodily joints. As described in section 3, we take twelve micro-poses for measuring the feature vector and the information is converted to symbols which represent a pose of trained image, and finally we can understand the meaning of the temporal gesture with code matrix as the code book of the Bag of words.

In the paper, we select 11 gestures of upper body as the recognition subject and those gestures are composed with 24 poses. To segment the micro-poses from the poses, we insert real-time training data more than 20 times for each pose. Fig.4 (a, b) show the procedure of the training.

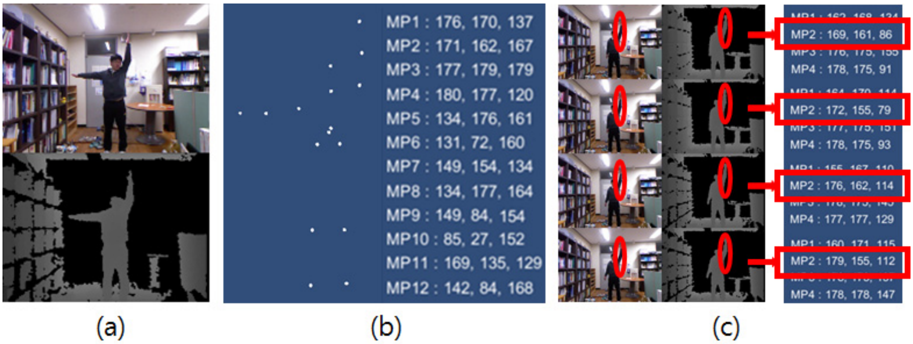
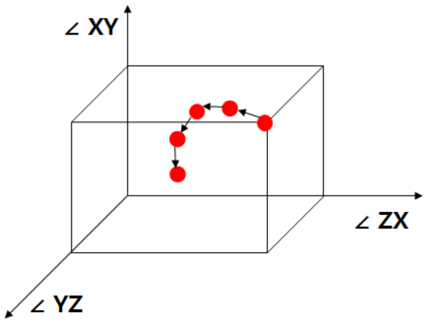


Fig. 4. Examples of micro-pose training scene ((a) color and depth image, (b) an example of joint feature vectors for a micro-pose, (c) an example of micro-poses; 'lift left hand.' The numbers represent the angle between joints.)

In the training process, the user can watch his/her own pose through a monitor and if he/she feels any discomfort on the pose he/she can change the pose. If the user makes an appropriate pose the system capture the data and it flush the twelve micro-poses into a file. Actually the micro-poses means angle difference among three specific joints.

Fig.4(c) shows an example of extracting a micro-pose from several input images. We use some different images for a micro-pose since real 3D position of bodily joints for the pose is very rugged even though the same person tries to make the same pose. Therefore we use multiple images as the training image.



∠ XY +	0	1	0	1	1
∠ XY -	0	0	1	0	0
∠ YZ +	0	0	0	0	0
∠ YZ -	0	1	0	1	1
∠ ZX +	0	0	0	0	0
∠ ZX -	0	0	1	0	1
N	1	0	0	0	0

(a) (b)

Fig. 5. Code matrix generation ((a) variation of the three dimensional angle values from a micro-pose, (b) code matrix generation using variation)

Fig.5 shows code matrix generation using variation of the three dimensional angle values. In the Fig.5(a), when we learn some micro-poses if angle is increased in the X-Y plane, ∠XY is increased one degree in the Fig.5(b).

The feature vectors of the micro-pose are consisted of angle between joints. We set the number of pose and micro-pose according to when is to recognize gesture.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14		T
∠XY +	0	21	0	20	0	20	0	0	21	0	20	0	0	0	0		0
∠XY -	0	0	0	0	19	0	0	0	0	19	0	0	19	0	0		0
∠YZ +	0	20	2	19	20	17	0	0	20	20	17	0	20	0	0		0
∠YZ -	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.....	0
∠ZX +	0	19	0	20	17	19	0	0	19	17	19	0	17	0	0		0
∠ZX -	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
N	22	1	18	1	1	2	22	22	1	1	2	0	1	22	22		22

Learned code matrix as the micro-poses of gesture A

Learned code matrix as the micro-poses of gesture B

∠XY +	1	0	1	0	1
∠XY -	0	0	0	1	0
∠YZ +	1	0	1	1	1
∠YZ -	0	0	0	0	0
∠ZX +	1	0	1	1	1
∠ZX -	0	0	0	0	0
N	0	1	0	0	0

User's code matrix as the micro-pose of a gesture

Fig. 6. Example of learned code matrix as the micro-poses

So, in this paper, we define the ten body gestures. And then, in order to generate micro-pose code matrix, we executed twenty-two learning by each micro-pose. The method which to learn pose is as follow Fig.6. In the Fig.6, the learned code matrix is compared with learned micro-pose's code matrix of gestures by using the Euclidean Distance.

To measure similarity using the Euclidean Distance, code matrixes are compared by applying as follow equation 1. In equation 1, n is number of valid column and m is number of valid row. And we estimate most similar to learned code matrix when the code matrix has maximum values by calculating similarity of the distance between the values.

$$\sum_{i=1}^{12}\sum_{j=1}^{n\times m}(Value_{Learned\ cm} \times Value_{User's\ cm}) \quad (1)$$

4 Conclusion

In this paper, we propose a unique algorithm which can identify the meaning of sequential bodily motion as a gesture by estimating deformed shapes of joints.

We call that micro-pose. The contribution of the approach is that we can reduce degree of freedom of articulated motions of a whole body since the variation of the pose is decomposed into small piece of joint combination. And the micro-poses are converted into symbols which are suitable for making use of code matrix as the code book of the Bag-of-Words.

The weak point of our approach is that gesture model is strictly limited to some gestures which have big variation of body shape. We think this defect comes from manual classification of gestures when we make the code book of model gesture. In the near future, we can make up the weakness by adopting automatic classification of model by using a clustering algorithm based on similarity measurement.

Acknowledgment. "This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1014) supervised by the NIPA(National IT Industry Promotion Agency)".

References

1. Ju, S.X., Black, M.J., Yaccob, Y.: Cardboard people: A parameterized model of articulated image motion. In: FG (1996)
2. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. IJCV 61(1), 55–79 (2005)
3. Wang, Y., Tran, D., Liao, Z.: Learning hierarchical poselets for human parsing. In: CVPR 2011, pp. 1705–1712 (2011)
4. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: ICCV, vol. 1, pp. 604–610 (2005)
5. Shotton, J., Fitzgibbon, A.W., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time Human Pose Recognition in Parts From Single Depth Images. In: CVPR (2011)