2016 16th International Conference on Control, Automation and Systems (ICCAS 2016)
Oct. 16-19, 2016 in HICO, Gyeongju, Korea

Continuous gesture recognition by using gesture spotting

Daeha Lee[1*], Hosub Yoon[2], Jaehong Kim[3]
[1,2]Intelligent Cognitive Technology Research Department, ETRI Daejeon, Korea
(bigsum, yoonhs, jhkim504)@etri.re.kr

Abstract: A gesture is not performed in only one action but a combination of continuous actions. It is very important to know the start and end of a gesture for accurate gesture recognition. In this paper, to extract a meaningful gesture portion in an online situation, we introduce a method that can distinguish the start and end of a gesture. Then, we describe the method of recognizing an extracted gesture.

Keywords: Gesture recognition, End-to-end detection

1. INTRODUCTION

As gestures are means of non-verbal communication, they are commonly used communication method in addition to the voice [1]. A gesture can be composed of a fixed single motion, but they are usually composed of many continuous movements [2]. As in speech recognition, in the gesture recognition, it is important to know the start and end points of a meaningful signal. If an observer starts to watch a gesture that has already begun, it would be impossible to correctly understand the gesture's meaning. For an online situation, to perform smooth gesture recognition, we detect the start and end point of a gesture, collect the meaningful target gesture, and finally identify it.

The remainder of this paper is organized as follows. In section II, we discuss how to distinguish between the start and end of the gesture. In section III, the proposed method is introduced. Next, experimental results are presented in section IV. Finally, in section V, we provide some concluding remarks and directions for future work.

2. END-POINT-DETECTION OF GESTURES

In speech recognition, using the end point detection method, the start and end points of a speech signal are obtained. By comparing the magnitude of the energy of the speech signal, a point that exceeds the threshold value is the beginning of the speech, and a point equal to or less than the threshold value is the end point of the speech, and then the portion between the start and end points of the speech is regard as a meaningful speech signal [3-4].

In gesture recognition, it is a very important to extract meaningful gesture parts from continuously performed movements. In speech recognition, by comparing the magnitude of the audio signal, end points of speech are obtained; to extract the end point of a gesture, a special action needs to be used to detect the end point of a gesture.

In this paper, we use the "Stand" posture as a special action for end point detection of a gesture. The "Stand" posture is easy to detect, and it is typically used when a gesture begins and ends, so we attempt to find the end points of gestures by using this action. Figure 1 shows an example of a continuous gesture, we know that there is a "Stand" posture at start and end of the gesture.

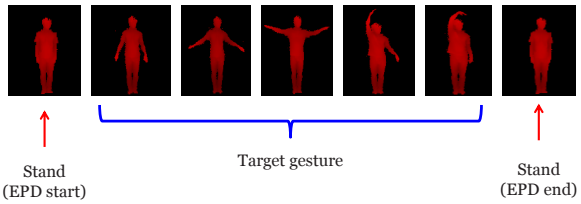

Fig. 1. Continuous gesture frames which have the "Stand" posture in the start and end frames.

Recognition of the "Stand" posture is carried out by utilizing skeletal joint information, and this process is illustrated in Fig. 2. To recognize the "Stand" posture, we use a combination of joint information. For example, in the "Stand" posture, the lengths of both arms are the same and both arms are straight and pointing downward. Also, both legs are straight.

- Stand
  <1> compare arm and spine dist
  Dist_leftShoulerToleftWrist,
  Dist_rightShoulderTorightWrist,
  Dist_neckToBase
  <2> compare arm angle
  Angle_leftShoulderToleftElbow
  Angle_rightShoulderTorightElbow
  Angle_leftElbowToleftWrist
  Angle_rightElbowTorightWrist

Stand position = <a> && <b>

Fig. 2. Skeletal joint information for the "Stand" posture.

3. PROPOSED METHOD

In this study, we developed an online recognition process for gestures that are used in interacting with a robot. Table 1 shows the related gestures.

978-89-93215-12-0/16/$31.00 ⓒICROS

Table 1. List of gestures

Gesture	Snapshot	Gesture	Snapshot
Up		Down	
Move backward		Move forward	
Move right		Move left	
Welcome		Agree	
Disagree		Love	

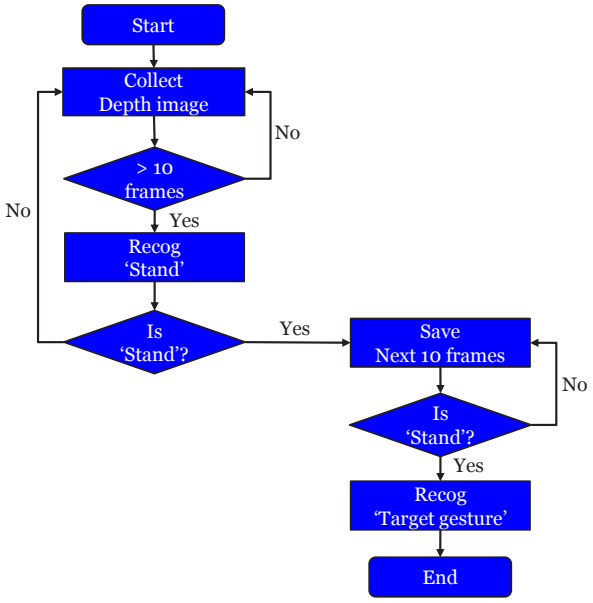

Fig. 3. Flow chart of gesture recognition.

We collect 10 frames and perform the "Stand" posture recognition process for each of the collected 10 frames. If 70% of the 10 frames are recognized as showing the "Stand" posture, those 10 frames are confirmed as a "Stand" block. Once recognized by the "Stand" block, it is designated as the start point for a meaningful target gesture, and after the start point, every 10 frames will be saved until the next "Stand" block is recognized. When the second "Stand" block is recognized, the saving task of target gesture is stopped and the recognition process on the saved target gesture will be performed. Figure 3 shows a flow chart of the gesture recognition process.

The recognition task for the target gesture is performed using motion history images (MHIs) [5] and the HOG algorithm [6]. We obtain the front / right/ left MHIs from depth gesture images and apply the HOG algorithm to each MHI. In Fig. 4, the target gesture conversion process is shown.

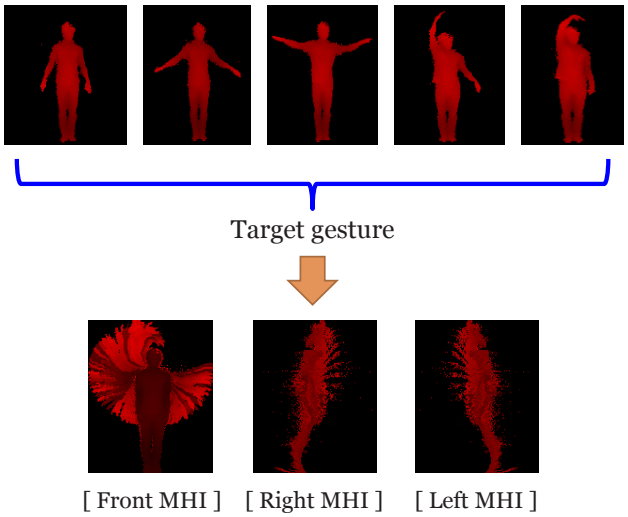

Fig. 4. Conversion process for target gesture.

4. EXPERIMENT & RESULTS

Based on the gestures defined in Table 1, we collected 1,000 gestures from 20 people. Among the total gestures, 500 gestures were used for the learning process, and the other gestures were used for the testing process. The gesture collection and recognition tasks were carried out using Microsoft's Kinect 2 device. Table 2 shows the results. The recognition rate achieved with a combination of front and side MHI was much better.

Table 2. Gesture recognition rates

Gesture Type	Accuracy (%)		
	Front MHI	Front / Right MHI	Front / Right / Left MHI
Up	92.5	97.5	97.5
Down	94	98	98.25
Move forward	91.75	92	93.5
Move backward	93.5	96.75	98.25
Move left	100	100	100
Move right	99.25	100	100
Welcome	98.5	98.75	98.25

Agree	97	96.75	97
Disagree	98.25	99.25	99.75
Love	96.75	97.75	97.75
Total	96.15	97.68	98.03

5. CONCLUSION

Gesture is a useful representation method not only among people but also between humans and animals. Gestures also have the potential to be a good communication tool between humans and robots. In this paper, we presented a method to enable effective online gesture recognition. Using the "Stand" posture, we detect the start and end points of a meaningful target gesture, and we perform the recognition process using front/right/left MHIs; thus, the gesture recognition result is improved. In the future, we will apply the proposed method to various fields.

ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MSIP/KEIT. [10041610, The development of automatic user information(identification, behavior, location) extraction and recognition technology based on perception sensor network(PSN) under real environment for intelligent robot]

REFERENCES

[1] Burgoon J., Buller D. "Nonverbal communication: The unspoken dialogue," New York: Harper & Row, 1989.

[2] Ju Yong Chang and Seung Woo Nam, "Fast Random-Forest-Based Human Pose Estimation Using a Multi-scale and Cascade Approach," ETRI Journal, vol. 35, no. 6, Dec. 2013.

[3] Lingyun Gu, Stephen. Z, "A new robust algorithm for isolated word endpoint detection," ICASSP Conf. Proc., 2002.

[4] Jia-lin Shen, T., "Robust entropy-based endpoint detection for speech recognition in noisy environments," Spoken Language Processing Conf., 1998

[5] A.F. Bobick, "The recognition of human movement using temporal templates," PAMI, vol. 23, 2001.

[6] N.Dalal, "Histograms of oriented gradients for human detection," CVPR, vol1, 2005.