

Detection and Recognition of Hand Gesture for Wearable Applications in IoMTW

Anna YANG*, Sung Moon CHUN**, Jae-Gon KIM*

* School of Electronics and Information Engineering, Korea Aerospace University, Korea

**Insignal, Korea

{nayang, jgkim}@kau.ac.kr, smchum@insignal.co.kr

Abstract—To support an efficient media consumption in a wearable device and IoT (Internet of Things) environment, the standardization of IoMTW (Internet of Media-Things and Wearables) is in the progress in MPEG (Moving Picture Experts Group). In this paper, we present a hand gesture detection and recognition algorithm to generate hand gesture-based commands for controlling the media consumption in smart glasses. In the proposed method, we use depth map and color image together to extract more accurate hand contour. We are going to present representation of the detected hand contour based on Bézier curve as metadata to provide an interoperable interface between a detection module and a recognition module. In a recognition module, the detected hand contour is reconstructed by parsing the delivered metadata. In the proposed recognition method, a set of hand gestures featured with diverse combination of open fingers and rotational angles can be recognized with quite stable performance in the proposed method. Finally, the recognized hand gesture is mapped into one of the pre-defined gesture commands.

Keywords— Internet of Media-Things and Wearable (IoMTW), Wearable, Smart Glasses, Hand Gesture Recognition

I. INTRODUCTION

Recently, the standardization of Internet of Media Things and Wearables (IoMTW) is undergoing in MPEG to support an efficient media consumption in a wearable device and IoT (Internet of Things) environment [1]. The normative parts of the IoMTW are mainly the specification of API (Application Programming Interface) and the representation of information to be delivered over the interfaces between media things including wearable devices and processing units in media-centric IoT and wearable applications. On the other hand, as hand gestures are attracting more attention as NUI (Natural User Interface) of wearable devices such as smart glasses, efficient detection and recognition of hand gestures is strongly required [2].

In this paper, we propose a hand gesture detection and recognition method for controlling the media consumption and/or controlling wearable device itself. In addition, the description of hand gesture is presented to support interoperable interfaces in the context of IoMTW.

The proposed method combines depth map and color image acquired from an incoming stereo image, and represents the detected hand contour based on a Bézier curve as metadata. The

detected hand contour delivered as a form of metadata is reconstructed in a recognition module. The reconstructed hand contour is recognized as one of the predefined set of hand gestures in a given application. Diverse types of hand gestures with different open fingers and angles can be identified mainly based on convexity defects of hand contour. A post processing is applied on the recognition results of consecutive frames in temporal domain to obtain more reliable results.

In Section II, we present gesture-based wearable application scenarios considered in IoMTW, and a gesture detection method is presented in Section III. A method of hand gestures description using Bézier curve is presented in Section IV, and a hand gesture recognition method using the described hand contour is given in Section V. Finally, the conclusions are presented in Section VI.

II. GESTURE-BASED APPLICATION SCENARIO

In IoMTW, gesture-based smart glasses applications are considered as one of the use cases [2]. When a user wear smart glasses, hand gestures are becoming an important user interface that allows a user to use both hands without operating an input device such as keyboard or mouse, etc. [3]. In order for a user to efficiently control applications for media consumption in a wearable device such as smart glasses, the hand gesture recognition should be supported in an IoMTW platform.

As shown in Fig. 1, user (user), IoMT (Internet of Media Thing), and processing unit (PU) are key functional components of IoMTW standard. In order to support various types of hand gesture-based use cases including the control of media consumption in a wearable device and/or device itself, hand gestures which include static ones or dynamic ones such as a trajectory of hand motion should be recognized into a user invoked command.

As shown in Fig. 1, the incoming stereo image sequence acquisitioned by the stereo camera mounted on the smart glasses is transmitted to the PU, in which user's gesture is detected and recognized (① in Fig. 1). We assume that gesture detection and gesture recognition more likely to be performed in separate PUs. These are the cases when 1) the detected hand contour or trajectory can be recognized as various types of gesture-based commands depending on the given application, and/or 2) the computational power of the PU may not be

sufficient to perform both processing of detection and recognition [4], [5].

Under this assumption, we need to specify a normative interface between both PUs performing detection and recognition, respectively. In particular, we focus on the representation of the detection results of hand contour to be delivered to the recognition module (② in Fig. 1). In other words, the hand region is detected and its contour is described as metadata in the hand detection module. Then, in the hand gesture recognition module, the detected hand contour is reconstructed by parsing the metadata delivered from the detection module, and the reconstructed contour is mapped into a user command by recognition. The commands are predefined many different ways according to the given application. The generated command controls the wearable device and/or the application SW of the wearable device (③ in Fig. 1). In this way, the use case of an efficient media consumption in a wearable device is implemented based on hand gestures.

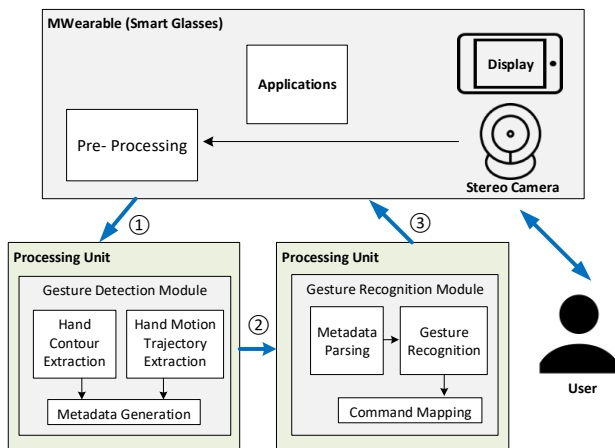


Figure 1. Use case of hand gesture based wearable application

III. HAND GESTURE DETECTION

Fig. 2 illustrates the overall procedure of hand gesture detection. A depth map is generated by stereo matching from the incoming stereo images. Due to the nature of the smart glasses, it is assumed that the user's hand is given at a certain distance range (30cm ~ 50cm) from the camera. Based on this assumption, the rough hand region is simply obtained by thresholding the value of the depth map. The detected rough hand region is refined by using color information and morphological filtering, which results in more accurate hand contour.

IV. HAND GESTURE DESCRIPTION

In order to deliver the detected hand contour in an interoperable way, we present an XML schema for describing hand contour based on Bézier curve. As shown in Fig. 3, a hand contour consists of a set of consecutive curves, and a single curve is represented by using nth-order Bézier curve given by Eq. (1). The nth-order Bézier curve is expressed by $(n + 1)$

points, which consist of a starting point, an end point, and a few control points that determine the curvature on the curve. In this paper, we use a cubic Bézier curve with the order three, so there are two control points in a curve.

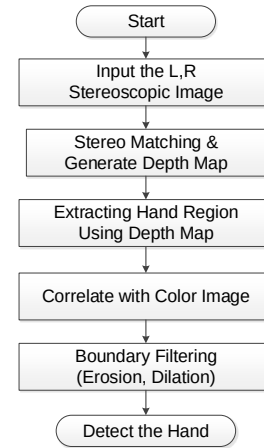


Figure 2. Overall procedure of hand gesture detection

$$Bezier(n,t) = f(t) = \sum_{i=0}^n \binom{n}{i} \cdot (1-t)^{n-i} \cdot t^i \quad (1)$$

An algorithm for representing hand contour using a Bézier curve is as follows. At the beginning, the entire hand contour is represented by a single Bézier curve, if the fitting error using a Bézier curve is larger than the threshold (max_error), and then the given contour is divided into two. This contour dividing process is repeated until the required accuracy is met.

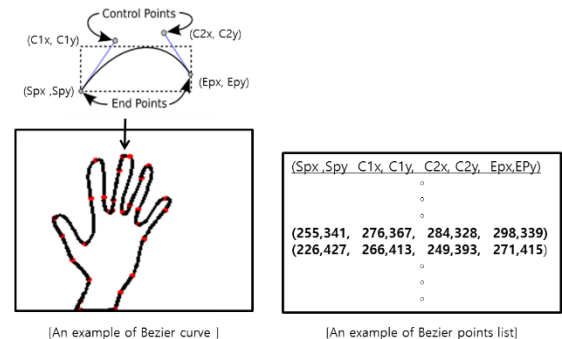


Figure 3. An example of Bézier curve

Fig. 4 shows an example of the result of this process. The given hand contour consists of a set of curve each of which is expressed by four Bézier points in the case of cubic Bézier curve. For example, four Bézier points are shown in Fig. 3 (starting point, two control points, and end point are given as (Spx, Spx), (C1x, C1y), (C2x, C2y), and (Epx, Epy), respectively).

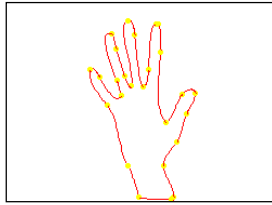


Figure 4. Result of hand gesture representation

The list of Bézier points representing a set of consecutive Bézier curves corresponding to the detected hand contour is described as XML-formatted metadata based on the hand gesture schema. The detected hand contour can be reconstructed by parsing the metadata in the recognition module. Fig. 5 shows an example of the reconstructed hand contour from the metadata.

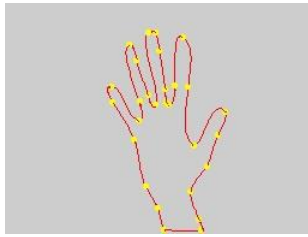


Figure 5. Reconstructed hand contour by parsing metadata

V. HAND GESTURE RECOGNITION

Fig. 6 shows the overall procedure of hand gesture recognition. Delivered metadata describing the detected contour is parsed into a list of Bézier points from which the detected contour is reconstructed. First, in order to recognize the hand gesture from the reconstructed hand contour, as an initial step, we generate a reference mask from the contour of open palm, which gives initial values of features that are used as reference values for the recognition step such as convexity defects, degree of rotation, and gravity center. Using convex hall, we can obtain some defects which allows us to figure out the position of each finger, and the gradient moment is used to obtain the center of gravity and degree of rotation.

The gradient of moment and the degree of rotation are obtained from the palm region by using Eg. (2) and Eg. (3), respectively. In order to obtain palm region, the finger region is removed from the hand region by morphological operations.

$$\mu_{p,q} = \sum_{(x,y) \in R} (x - \bar{x})^p (y - \bar{y})^q \quad (2)$$

$$\theta = \frac{1}{2} \tan^{-1} \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \quad (3)$$

In the recognition step, the reference mask is superimposed on the reconstructed hand contour for each frame. In this way, the finger state of open or close is identified by comparing the initial defects and the current defects. In the recognition algorithm, in which states of each fingers as well as the degree

of rotation are identified based on the initial features of the reference mask.

In order to figure out whether the finger is opened or closed, an ellipse is drawn in the area of the finger contour, and if there are more pixels in the finger region than the threshold value inside of the ellipse, it is regarded that the corresponding finger is spread. It can be also recognized which fingers are spread by checking the defects inside of ellipse. Finally, the gesture of OK sign can be recognized by checking the existence of contours inside of the hand contour.

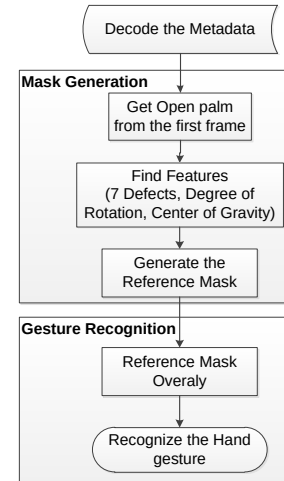


Figure 6. Overall procedure of hand gesture recognition

With the final result of the recognition, the command corresponding to the recognized hand gesture is generated by mapping the recognized hand gesture into one of the predefined set of commands.

VI. EXPERIMENTAL RESULTS

In the experiment, a set of hand gesture to be recognized is given in Fig. 7, which include diverse types of hand gesture with different combination of open fingers. Using the seven defect points, it is possible to recognize which fingers are open as well as the number of open fingers. In addition, the gesture of OK sign can be recognized by checking the existence of inside contour. Some examples of recognition results are shown in Fig. 8. The left and right images in Fig. 8 are the stereo images acquired by the stereo camera, and the centered small image shows recognition result. The results for the same number of fingers but different postures are shown in Fig. 8 (a). Fig. 8 (b) is the recognition result for the OK sign. And the recognition result depending on the degree of rotation of the hand is shown in Fig. 8 (c).

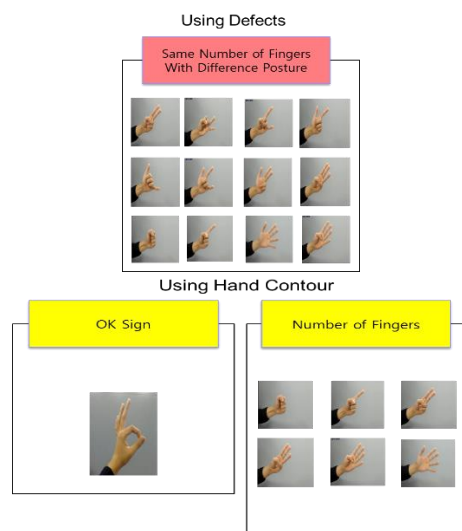


Figure 7. A set of hand gesture to be recognized

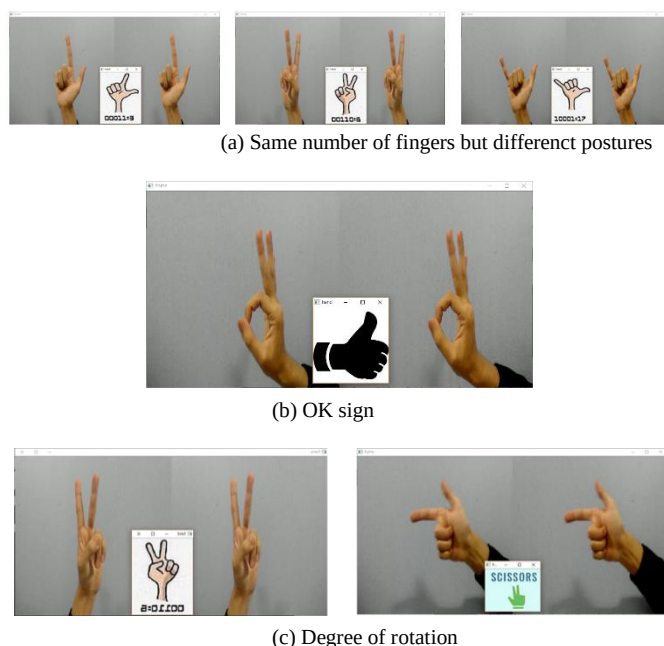


Figure 8. Some results of hand gesture recognition

VII. CONCLUSIONS

In this paper, we presented a hand gesture detection and recognition method for gesture-based smart glass applications in IoMTW. In addition, we presented the method of representation of hand contours using Bézier curves in the context of IoMTW. In the way, the system configuration of deploying the detection module and recognition module are well suited to IoMTW environment with interoperable interfaces between both modules. Experimental results showed that the proposed methods of detection and recognition of hand gestures could be effectively applied to wearable applications such as smart glasses in real time. As further work, we are going to developing a method to accommodate dynamic hand gestures considering hand motion trajectory.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by Korean Government (MSIP) (R0127-15-1015).

REFERENCES

- [1] "Call for Proposal on Media-Things and Wearables," ISO/IEC JTC1/SC29/WG11 W16535, Chengdu, China, Oct 2016.
- [2] M. Mitrea, S.-K. Kim, and S. Chun "Use cases for Internet of Media-Things and Wearables," ISO/IEC JTC1/SC29/WG11 N16533, Chengdu, China, Oct 2016.
- [3] S. Chun, "International standard of Wearable MPEG: Multimedia Communication and Gesture Recognition," Broadcasting and Media Magazine, vol. 20, no. 2, April 2015.
- [4] A. Yang, D. H. Park, S. Chun, and J. G. Kim, "Detection of Hand Gesture and its Description for Wearable Applications in IoMTW," In Proc. 2016 Summer Conference on The Korea Institute of Broadcast and Media Engineers, Jeju, Jul 2016.
- [5] A. Yang, S. Chun, H. Ko, and J. G. Kim, "Hand Gesture Description for Wearable Applications in IoMTW," ISO/IEC JTC1/SC29/WG11 m38526, Geneva, CH, May 2016.



Anna Yang received the B. S. degree from Korea Aerospace University, Korea, in 2014. She is currently working toward the M.S degree in the Department of Electronics and Information Engineering, Korea Aerospace University, Goyang-city, Korea. Her current research interests include video coding, internet of media things and wearables, and the wearable applications.



Sung Moon Chun received the B. S. degree from Sungkyunkwan University, Korea, in 1990. From 1990 to 2000, he was with Hyundai Electronics, where he was involved in the development of HDTV, MPEG-2/4 standards and related video codec technologies. From 2002 to 2012, he was with ECT Inc, where he was involved in the development of Semiconductor and related stereo video technologies. he is currently CTO in Insignal Inc. His research interests include video compression, virtual reality, wearable application.



Jae-Gon Kim received the B.S. degree in electronics engineering from Kyungpook National University, Daegu, Korea, in 1990, the M.S. and Ph.D degrees in electronical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1992 and 2005, respectively. From 1992 to 2007, he was with Electronics and Telecommunications Research Institute (ETRI), where he was involved in the development of digital broadcasting media services, MPEG-7/7/21 standards and related applications, and convergence media technologies. From 2001 to 2002, he was a Staff Associate at the Department of Electrical Engineering, Columbia University, New York, USA. Since 2007, he has been with the Korea Aerospace University, Goyang-si, Gyeonggi-do, Korea, where he is currently a professor in the School of Electronics and Information Engineering. His research interests include digital video coding, video signaling processing, digital broadcasting media, and multimedia applications.

[6]