TABLE III
RESULTS ON THE COMPARISON BETWEEN MPEG-7 CSS AND BÉZIER CURVE BASED DESCRIPTION

Items	Bézier curve based descriptor	MPEG-7 CSS descriptor	Comment
Applicability	Contour of the object Moving trajectory of the object	Contour of Object	
Scalability	Scalable in terms of description size and contour accuracy	Not available	Possible to trade off contour accuracy and description size
Reconstruction	Available	Not available	
Computational Performance	Average ~ 5msec	Average ~ 20msec	Computational time for generating descriptor from hand contour
Description Size	Hand contour: Average 95Byte Hand trajectory: Average 36Byte	Average 15Bytes	Both cases are very light weight
Main Applications	General contour description, transmission and reconstruction	Shape similarity matching and indexing	

Using convex hall, we can obtain some defects which allow 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 $\mu_{p,q}$ and the degree of rotation θ 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. x and y mean a coordinate of pixels, $\bar{x}$ and $\bar{y}$ mean a mean of x and y in the palm region, respectively.

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

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

In the recognition step, the reference mask is superimposed on the reconstructed hand contour for each frame depending on the rotation of degree. In this way, each finger's state of open or close is identified by comparing the initial defects and the current defects. Fig. 9 shows the details on the recognition algorithm, in which states of each finger as well as the degree of rotation are identified based on the initial features of the reference mask.

The details on the algorithm for hand gesture recognition are given in Fig. 9. When the reconstructed hand contour is inputted for recognition of hand gesture. First of all, 7 defects which are used for recognition of open finger position are founded by using convexity defect algorithm. To match between the reference mask and reconstructed hand contour, after then, the degree of rotation of reconstructed hand contour should be calculated by using the gradient of moment. In this way, the reference mask is able to be mapped into the hand gesture contour according to its rotation.

In order to figure out whether the finger's state is open or close, an ellipse is drawn in the region finger, 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 finger is spread by checking the defects inside of ellipse. Finally, the gesture of 'OK' sign can be recognized by checking the existence of a closed contour inside of the hand contour.

As the result of the recognition, the command corresponding to the recognized hand gesture is generated by mapping the recognized hand gesture into one of the set of commands predefined in a given application.

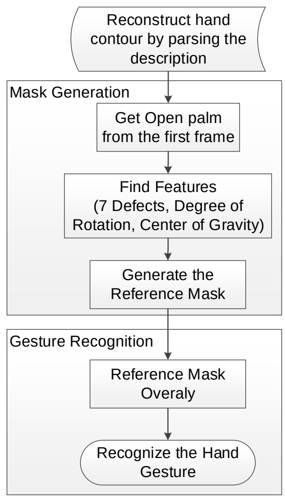


Fig. 8. Overall procedure of hand gesture recognition from reconstructed hand contour by parsing the delivered metadata

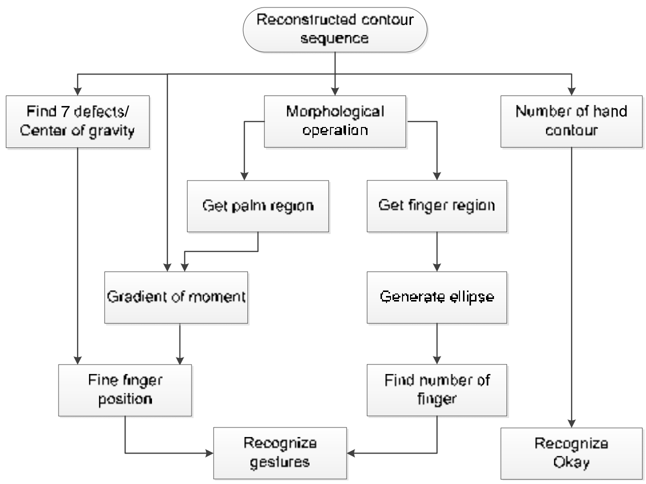
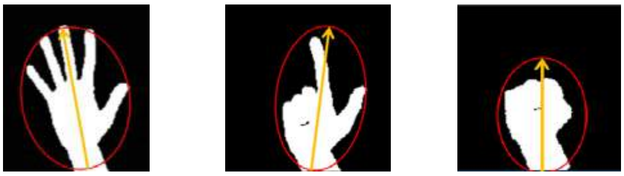


Fig. 9. Detains on the procedure of the hand gesture recognition

By the way, the rotation of reference mask may be incorrectly calculated by the center of gravity when few fingers are opened or no finger is opened, which results in directionally biased form as shown in Fig. 10 (b) and (c). To recognize a rotated hand correctly by overcoming this problem, the rotational degree of reference mask should be revised according to the rotation of given hand contour.

To address this issue, the rotational angle of the reference mask is obtained more accurately by using Eq. (4). In other words, we consider an additional factor of palm rotational angle as well as both rotational angles of the incoming hand gesture and the reference mask. We use a morphological method to obtain either palm region or finger regions only. Then, the palm region is used for the calculation of rotation, and mask overlay with the consideration of the rotation.



(a) (b) (c)

(a) A reference mask superimposed on the open hand
(b) A reference mask superimposed on the biased hand
(c) A reference mask superimposed the rock position hand

Fig. 10. An example of reference mask rotated depending on hand poses

As shown in Eg. (4), the value of R_{ref_mask}, the angle of reference mask that should be rotated when it is superimposed on the hand contour is compensated with the difference of rotational angle of palm and gesture.

$$R_{ref\_mask} = \left(A_{gesture} - A_{mask} + \left(A_{palm} - A_{gesture}\right)\right) \cdot \frac{\pi}{180} \quad (4)$$

$A_{gesture}$: rotational angle of the hand gesture in the current frame

A_{mask} : rotational angle of the reference mask

A_{palm} : rotational angle of finger removed palm in the current frame

The obtained finger region is used for identifying whether a finger is open or not. In other words, we apply an ellipse to each finger regions, and determine the existence of finger according to the number of pixel lies inside the region overplayed ellipse. If the number of pixel is larger than threshold, then the existence of finger is true, which is marked '1'.

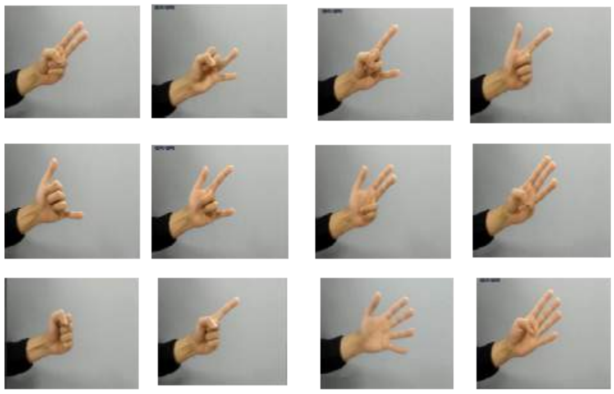
Finally, a post processing is applied on the recognition results of consecutive frames in temporal domain to reduce recognition error in each frame. The recognition error is occurred by not only in the process of changing hand gestures but also by fine shaking of hand. By accumulating the result of each frame, more reliable recognition result can be obtained for a given duration in the temporal post processing.

VI. Experimental results

In the experiment, a set of hand gestures to be recognized is given in Fig. 11, which includes diverse types of hand gesture with different combination of open fingers, different hand postures with the same number of fingers, and 'OK' sign. Using the seven defect points, it is possible to recognize which fingers are open as well as the number of open fingers. The proposed method uses the number of finger contour to figure out the number of open finger, and the position of open finger is recognized by using ellipse and convexity defects.

Example results of recognition are as shown in Fig. 12. As a result of hand gesture recognition, the position of each finger is obtained as '1' (open) or '0' closed. The binary code representing the position of five fingers are compared with those of the predefined gestures. Finally, the gesture image corresponding to the matched gesture is displayed as a final result as shown in Fig. 13. In addition, the gesture of 'OK' sign can be recognized by checking the existence of a closed contour inside the contour.

Some examples cases of final recognition results are shown in Fig. 13. The left and right images are the stereo images acquired by the stereo camera, and the centered small image shows the final recognition result. The example results for the same number of fingers with different postures are shown in Fig. 13 (a). Fig. 13 (b) shows the recognition result for the 'OK' sign, and the recognition result depending on the degree of rotation of the hand is shown in Fig. 15 (c). In this way, any hand posture can be identified and mapped into the corresponding gesture based user's command in wearable applications

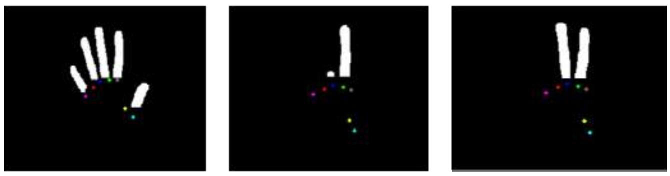


(a) Hand gestures with different postures of hand

(b) OK sign

(c) Hand gesture with different number of finger

Fig. 11. A set of hand gestures to be recognized in the experiment



(a) Recognition results of each finger's position

Finger	Bin
THUMB	1
INDEX	1
MIDDLE	1
RING	1
LITTLE	1

Finger	Bin
THUMB	0
INDEX	1
MIDDLE	0
RING	0
LITTLE	0

Finger	Bin
THUMB	0
INDEX	1
MIDDLE	1
RING	0
LITTLE	0

(b) Binary code related with (a) for representing each finger's position

Fig. 12. Example results of hand gesture recognition

In the experiments, to measure the recognition accuracy of the proposed method, we use a pre-recorded video in addition to real-time video capturing hand gesture. The experimental conditions are summarized in Table 4, and the recognition accuracy is measure by using Eq. (5).

$$Recog\_accuracy(\%) = \frac{num\_of\_accurately\_recognized\_frame}{num\_of\_total\_frame} \quad (5)$$

(a) Same number of fingers but differenct postures

(b) OK sign

(c) Gesture with different degree of rotations

Fig. 13. Some final results of hand gesture recognition

TABLE IV
EXPERIMENTAL CONDITIONS

PC	Processor: Inter® Core™ i7-6700 CPU @3.40 GHz
	Memory: 16.0GB
	OS: Windows 10, x64
Input stereo image sequence	Frame rate: 6 frames/sec Resolution: 1280 * 720 Number of gesture type: 12 types Number of Frame: 252 frames

Table 5 shows the performance of proposed method in terms of recognition accuracy measured by Eg. (5). The comparison results are for both cases of pre-recorded video and real-time video with and without the temporal post-processing.

Pre-recoded video are captured in an ideal environment without illumination change and no movement of hand location except hand gestures themselves. In such ideal environment, we obtained the accuracy of 97.2% in a recognition of finger numbers and 95.0% of recognition of finger position, respectively. However, in experiments on real-time video captured by the head mounted camera, illumination and location of hand are not stable due to a little motion of head. As a result, the recognition accuracies are decreased to 95.2% and 89.9%, respectively.

With the temporal post-processing, the recognition accuracy of the number of finger and finger position using pre-recorded video are 100.0% and 98.6%, respectively. In the case of real-time video, the accuracies are given by 97.0% and 93.0%, respectively. This results mean that the proposed temporal post processing significantly enhances the performance of recognition accuracy higher.

TABLE V
EXPERIMENTAL RESULTS OF RECOGNITION ACCURACY

Test condition	Recognition accuracy			
	Pre-recorded video		Real-time video	
	Finger number	Finger position	Finger number	Finger position
Without post processing	97.2	95.0	95.2	89.9
With post processing	100.0	98.6	97.0	93.0

VII. CONCLUSION

In this paper, we presented a hand gesture detection and recognition method for gesture-based smart glasses applications in the context of MPEG IoMT, which aims to support efficient media consumption in IoT and wearable environments. In addition, we presented a method of representation of hand contours using Bézier curves to provide an interoperable interface between processing units each of which perform gesture detection and gesture recognition, respectively, in an IoMT framework.

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. In addition, we have found that the Bézier curve based descriptor is more appropriate to support hand gesture based use cases than the MPEG-7 CSS and MPEG-U Part 2 based on the comparison analysis.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by Korean Government (MSIT) (R0127-15-1015). This paper is an extension of the ICACT 2017 conference paper.

REFERENCES

[1] M. Mitrea, *Working draft 2.0 of ISO/IEC 23093-1 IoMT Architecture*, ISO/IEC JTC1/SC29/WG11 N17094, Torino, July 2017.
[2] H. Cheng, L. Yang, Z. Liu, "Survey on 3D Hand Gesture Recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1659-1673, Sep. 2016.
[3] 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.
[4] A. Yang, S. Chun, H. Ko, J. G. Kim, *Hand Gesture Description for Wearable Applications in M-IoTW*, ISO/IEC JTC1/SC29/WG11 M38526, Geneva, Swiss, May. 2016.
[5] A. Yang, S. Chun, and J.-G. Kim, "Detection and Recognition of Hand Gesture for Wearable Applications in IoMTW," In *Proc. ICACT2017*, Feb. 2017, pp. 598 – 601.
[6] *Information technology – Rich media user interfaces – Part 2: Advanced user interaction (AUI) interfaces*, ISO/IEC 23007-2:2012, Oct. 2012.
[7] *Information technology – Multimedia content description interface – Part 3 Visual*, ISO/IEC 15938-3:2002, Oct. 2001.
[8] D. Zhang, G. Lu, "Evaluation of MPEG-7 shape descriptors against other descriptors," *Multimedia Systems*, vol. 9, no. 1, pp. 15-30, July 2003.

Anna Yang received the MS. degree from Korea Aerospace University, Korea, in 2017. She is currently working toward the Ph.D degree in the Department of Electronics and Information Engineering, Korea Aerospace University, Goyang-city, Korea. Her current research interests include IoT/wearable media applications, and MPEG standards.

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.

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, and wearable applications.