MPEG-U based Advanced User Interaction Interface System Using Hand Posture Recognition

Gukhee Han, Haechul Choi

* Department of Multimedia Engineering, Hanbat National University, Korea
guk_k@hanbat.ac.kr, choihc@hanbat.ac.kr

Abstract— **Hand posture recognition is an important technique to enable a natural and familiar interface in HCI (human computer interaction) field. In this paper, we introduce a hand posture recognition method by using a depth camera. Moreover, the hand posture recognition method is incorporated with MPEG-U based advanced user interaction (AUI) interface system, which can provide a natural interface on a variety of devices. The proposed method initially detects positions and lengths of all fingers opened and then it recognizes hand posture from pose of one or two hands and the number of fingers folded when user takes a gesture representing a pattern of AUI data format specified in the MPEG-U part 2. The AUI interface system represents user's hand posture as the compliant MPEG-U schema structure. Experimental results show performance of the hand posture recognition and it is verified that the AUI interface system is compatible with the MPEG-U standard.**

Keywords— **MPEG-U, Hand and finger recognition, Advanced User Interaction, User interface**

I. INTRODUCTION

Recently in the human computer interaction (HCI) field, many researches have been conducted on the interaction which gives and takes the communication for the interface between the human and the machine. The conventional interface methods include the switch based method using tools like a keyboard or a mouse and the method using pointing devices. But, these are not the methods as natural as the human communication. The natural interface method should consider the characteristics of the human communication system. Voice and gesture are the most common means of the human communication. According to [1], human acquires more than 80 % of the information through visual method, thus visual interface is one of the most familiar and convenient methods for human among various interface methods. To enable the visual interface between the human and the machine, a technology to recognize user's intent represented by gesture or posture of hands is needed. Moreover, a practical interaction system utilizing recognition results as the natural interface should be developed, which requires compatibility with a variety of machines.

There are various studies on the hand recognition. Many hand recognition methods segment hand region based on color information such as skin tone [2, 3, 4]. However, these methods dependent on color information may be sensitive to practical environment such as illumination change and the skin color of race. To detect hand region more efficiently, a specific glove or a wrist cover hand is utilized in [5, 6], the 3 dimensional hand model is used in [7], and 3 dimensional image data that are an image data of the 2 dimensional image data plus information on the depth are used in [8].

In order to recognize the hand's posture correctly, the fingers should also be recognized. There are various works in the literature. In [9], fingers are detected by using curvature between fingers obtained from the gradient of luminance. In [10], the convex hull based convexity defects are uses as the feature for the finger detection. In [11], the edge based finger phase model is introduced for the probabilistic matching. In [12], fingers are detected according to the connectivity of horizontal lines within a hand region.

For the purpose of providing a familiar and natural user interface, this study considers the hand posture as the natural interface. The robustness of the hand detection which copes with the various environmental changes is realized based on the 3 dimensional image data including the depth information [8]. The depth information can be obtained with 2 or more cameras or a depth sensor. Recently, there have been an increasing number of depth cameras available at commodity prices and also there is trend of releasing the products where the smart machine is equipped with depth camera. Thus, the usage of the depth camera can be a more practical solution than that of specific equipment such as the glove or the wrist cover hand. As for the finger detection, this study also uses the convex hull based convexity defects [10], which is robust to noise and requires relatively low computational complexity.

One of the points that should be considered in the user interface system is the interoperability among various machines. The conventional interface data format representing user's hand gesture or posture is differently specified on the customized software or even hardware platform, which may not be compatible with various machines. Thus, in the ISO/IEC JTC1/SC29/WG11 MPEG(Moving Picture Experts Group), the advance user interaction (AUI) interface data format is being standardized, and that standardization project is called MPEG-U part 2 AUI Interface [13]. The purpose of the MPAG-U part 2 is to support effective interoperability and interface method by providing the data format which has consistency defined as standard in the interface [14]. In the MPEG-U part 2, the above mentioned AUI interface data format is defined according to the XML based schema

structure. The design goals of XML emphasize simplicity, generality, and usability over the internet. Although the design of XML focuses on documents, it is widely used for the representation of arbitrary data structures, for example in web services. Accordingly, this study introduces a MPEG-U part 2 based AUI interface system that can be interoperable among different machines.

II. PROPOSED SYSTEM

The paper introduces a MPEG-U based AUI interface system that provides a natural interface using the hand posture. The hand posture is recognized with the depth camera, KINECT. Figure 1 shows the proposed system that consists of the two parts, the hand posture recognition and MPEG-U XML creation. The hand posture recognition part is further divided into the detection module and the recognition module. In the detection module, a hand region is roughly detected applying the hand detection algorithm provided in the Candescent NUI [8]. To minimize the size of the detected hand region, the position of the wrist and the end positions of fingers are located and the roughly detected hand region is refined with these positions. Next, the center point of the minimum hand region is calculated for the finger detection and the end point of the fingers are detected using the Convex Hull based Convexity Defects. The recognition department applies the algorithm which differentiates between the left hand and the right hand and among the thumb, index finger, middle finger, ring finger and the little finger and differentiate them based on the detected result, and judges the hand's posture and the fingers' spread status and recognize the final hand posture. Next, the MPEG-U XML creation part takes the posture which has the characteristic in accordance with the MPEG-U part2 standard and parses it into the XML document in accordance with the MPEG-U part2 standard in order to express the recognized hand posture into the interoperable data format.

A. Hand posture recognition part

1) Detection department: In the KINECT, the time of the infrared light being shot and coming back reflected is measured by the detection sensor and the depth information is acquired [15]. In order to use the KINECT, the 3D motion recognized middleware which enables the compatibility with the PC is needed. The middleware used mainly for the KINECT are KINECT SDK from Microsoft and the Prime sense OpenNI etc. Candescent NUI provides the open source which recognizes the hand and the fingers utilizing these middleware. Candescent NUI has the advantage of being able to detect the hand precisely and swiftly utilizing the 3 dimensional image data and is robust to the change of the complex background and illumination and the computational complexity is also very low. The hand and finger recognition part of the proposed system uses the open source of Candescent NUI [8] for the depth information input from the Kinect to detect the hand region as in the figure 2-(a). In order to detect the minimum hand region among these detected overall hand regions, the followings are performed.
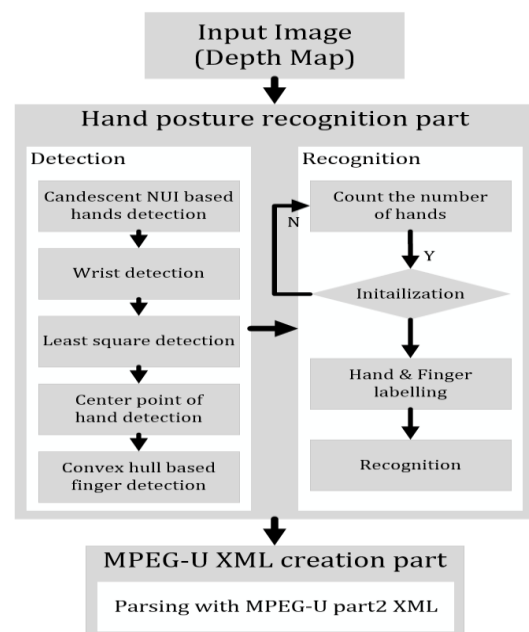

Figure 1. The proposed hand posture recognition and MPEG-U based XML creation

First, the hand region is fitted to a line in the sense of the minimum least square error [16]. Then, pairs of two intersection points between the outermost line of the hand and the normal line of the fitting line are found. When the variance of the distance between a pair of the two points is constant, the normal line through the pair of the two points is decided as the wrist line. Figure 2-(b) is a figure which indicates the point of the wrist by indicating one line which represents the overall hand region. After finding the wrist point, calculate the least square including the minimum hand region from the wrist point to the end part of the fingers. Figure 2-(c) is a figure which indicates the least square including the minimum hand region from the wrist point to the end point of the fingers. Next, find the center point of the minimum hand region by applying the distance transform algorithm [17]. The center point is necessary for the detection and recognition of the fingers afterwards and is often utilized as the position value of the hand in the MPEG-U AUI interface data format. Figure 2-(d) shows a figure which indicates the center point of the minimum hand region which applies the distance conversion algorithm. Next, detect the end point of the fingers using the Convex Hull base Convexity Defects [10]. This method has the advantage of assuming the correct fingers and obtaining the directional information on the corresponding fingers, but there is disadvantage that it needs necessarily additional post process procedure. The additional post process procedure is as follows. Assess the goodness of fit with the oval approximation model improved based on the oval morphological characteristic that the finger joint has. Based on the result above, detect the final end point of the fingers by using information on the vector angle difference between the end points of each finger and the center point. In figure 2-(e), the red dots are the finally detected end points of the fingers.
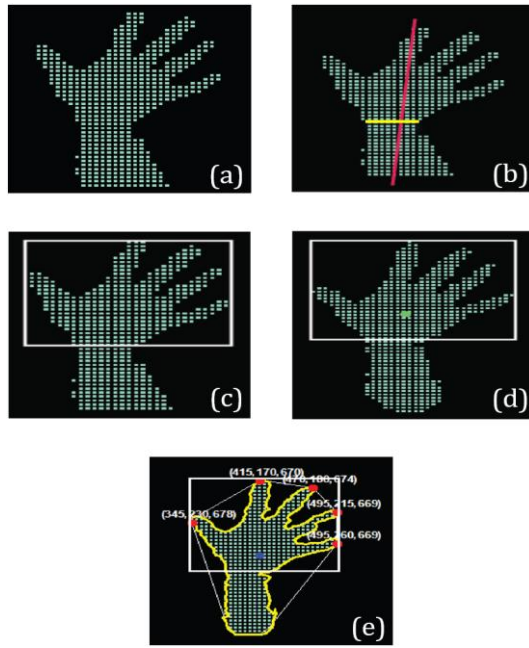
Figure 2. result of the hand detection using the Candescent NUI, (b) result of the wrist detection by applying the line fitting algorithm, (c) result of detecting the least square of hand region, (d) result of detecting the center of mass of the least square applying the distance transform algorithm, (e) result of finger detecting using the Convex Hull based Convexity Defects

2) Recognition department: The recognition department classifies with the right hand and left hand. And classifies with the thumb, index finger, middle finger, ring finger and the little finger. It is based on the detected hand and fingers. The classification of the hand is performed considering the situation where the minimum hand region is either one hand or both hands. If both hands, by the topology judgment, what is in the left is classified to be the left hand and what is in the right is classified to be the right hand. Finger classification applies the proposed finger skeleton classification algorithm as in the figure 3. First, the number of regions detected of the hand is calculated. When the number of hand region is two, the numbers of center points and end points of fingers are initialized to be two and ten, respectively. When the number of hand region is 1, the numbers of those points are initialized to be one and five, respectively. Next, a skeleton of the hand is made as follows. When the center point is $C(x_c,y_c)$ and the end points of each finger are $F_i(x_i,y_i)$, (i=0~4), the skeleton of each finger L_i will be defined as the line which connects the C and F_i. And the length is defined as Length_L_i. The two shortest skeletons among the Length_L_i of each skeleton is found and the thumb's skeleton candidate is found, and the possible thumb's skeleton candidate Ls_1, Ls_2 is called. The skeleton of which the value is lower than the critical value among the thumb's skeleton candidate is selected. Then it is called optimal thumb's skeleton. And when the value is lower than or greater than the critical value, it is called optimal thumb's skeleton by using the angle among the adjacent skeletons. When we call the angle among the adjacent skeletons of Ls_1 Angle_s_1 and the angle among the adjacent skeleton of Ls_2

Angle_S_2, the skeleton is classified which has the maximum angle. After classification the thumb's skeleton, the rest of the fingers in the order of most adjacent fingers to thumb is classified as the skeleton of the index finger, middle finger, ring finger and the little finger. After that, by using the classified finger's skeleton, if the end point of the finger lies on the corresponding skeleton or lies adjacent to the corresponding skeleton, is classified that end point of the finger as the end point of the finger of that corresponding skeleton. For the real time application, it is not created a new skeleton when inputting each time, and the detection department only is executed, and is compared with the acquired least square to the previous point of least square and is calculated the coefficient of the movement and rotation conversion and apply this to the skeleton created at the previous point and decrease the complexity for the creation of the skeletons. Next, the final hand posture is recognized by judging the hand posture and if each finger is spread.
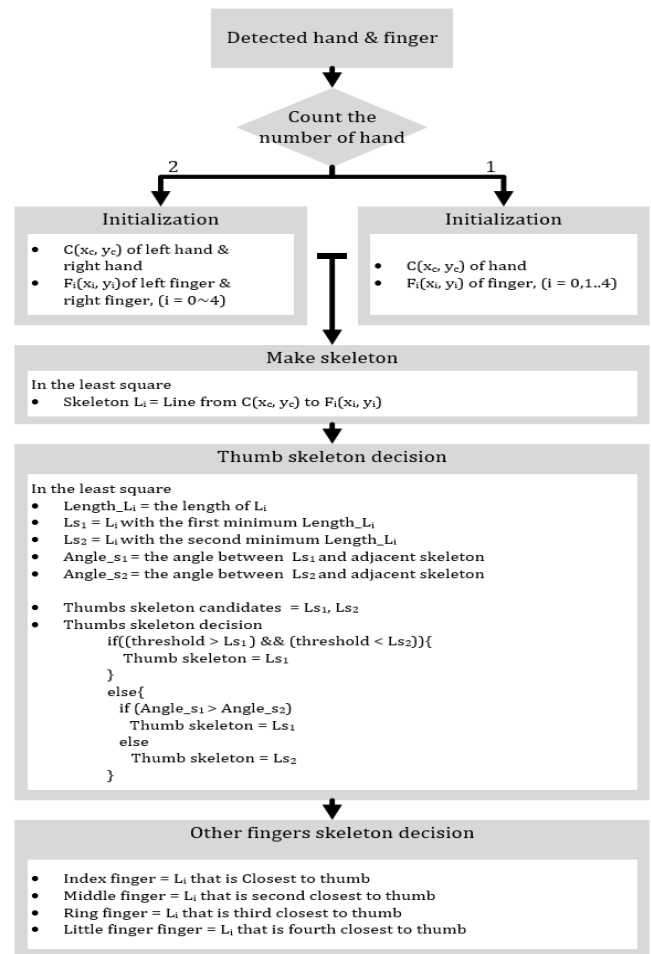

Figure 3. Proposed finger skeleton classification algorithm

B. MPEG-U XML creation part

The creation part of MPEG-U XML creates the XML document according to the MPEG-U Prt2 standard in accordance with the schema structure defined in the MPEG-U part 2. MPEG-U part2 supports the schema structure for all
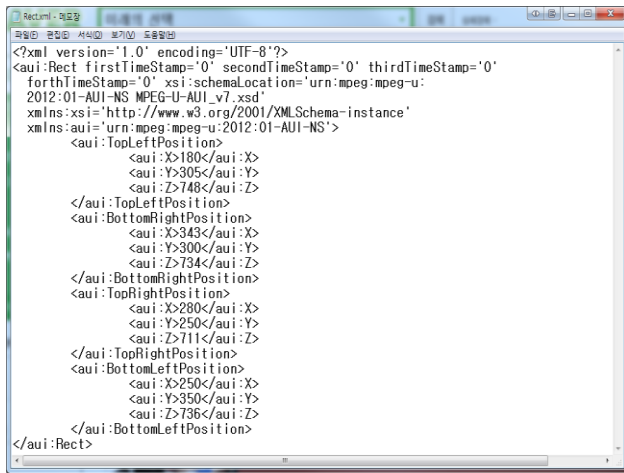
Figure 3. Rect XML of MPEG-U part2 geometric pattern parsed when a user takes the Rect posture

data type as XSD document. If user poses of scissors using his recognized hand and fingers, in the creation part of MPEG-U XML creates the XML document using the schema structure scissor type. The information of scissor type is included in the document and the parameters processed in the hand and finger recognition part are divided into the X, Y, Z coordinates in accordance with the MPEG-U part2 standard and is narrated. Figure 4 is the figure parsed in accordance with the MPEG-U part2 standard about the posture which has the characteristic which corresponds to the MPEG-U part2 standard.

III. EXPERIMENT

This chapter indicates the end points of the fingers using Convex Hull based Convexity Defects through an experiment and shows the result of recognition of the proposed finger classification algorithm. Also, it validates through experiments the points where the hand posture recognition is possible in various environments. Finally, it validates the suitability of the MPEG-U based advanced user interface system for the interconnection by experimenting the proposed system with MPEG-U part2 reference software.
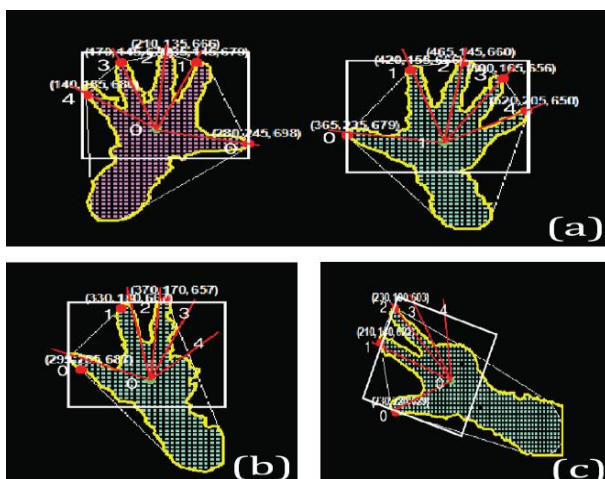

Figure 5. Hand posture recognition result of the proposed method (a) result when all the 10 fingers are spread out, (b) result when a few fingers are unfolded, (c) result under translation and rotation of hand

A. Proposed finger classification algorithm verification

The proposed system recognizes the hand as the left and right hand and the fingers as the thumb, index finger, middle finger, ring finger and the little finger. The posture is robustly recognized in the experiment when all the 10 fingers are spread or only a few of them are spread or when they move or rotate. Figure 5-(a) is the test result which shows the hand and fingers when all the 10 fingers are spread and figure 5-(b) is the test result that shows the hand and the fingers when only a few fingers are spread. Figure 5-(c) is the test result which classifies the hand and the fingers when the user takes the posture giving change of movement and rotation. In the figure, the 0 and 1 of the hand region indicates the left and right hand and the 0~4 in the adjacency of the fingers indicates each thumb, index finger, middle finger, ring finger and the little finger.

B. Hand posture recognition verification in a variety of environments

The proposed system uses the 3 dimensional image data and thus is very robust to the various environments. The Candescent NUI which detects the hand utilizing the 3 dimensional image data is robust to various environmental changes and the computational complexity is small. In the experiment, the system shows the performance of the proposed method in the background of which the color is similar to the skin color and the change of illumination. Figure 6-(a) is the figure that shows the hand recognition result recognized in the bright background of high brightness value and the figure 6-(b) is the figure that shows the hand recognition result recognized in the dark background of low brightness value. Figure 6-(c) is the figure that shows the hand recognition result recognized under the background having the similar color with the skin color. As shown in the figure, this proposed method is robust to the change of the environment and the illumination.

C. The MPEG-U based advanced user interaction interface system verification

It verifies the system proposed by experimenting the type about Geometric Pattern, Symbolic Pattern among the 6 different patterns defined in the MPEG-U part2. The types to be used in the experiment are the Rect of Geometric Pattern and the Scissors of Symbolic Pattern. The Rect of Geometric Pattern needs the data on the Top Left Position, Bottom Left Position, Top Right Position and the Bottom Right Position and the Scissors of Symbolic Pattern needs the size data. In the proposed system, in order to express the Rect and Scissors, in the Rect the classified thumbs and index fingers of both hands are used and in the Scissors, the thumb and index finger of one hand and the size of the hand region are used. Figure 7 is the figure which expresses the Rect and Scissors with the recognized hand posture and figure 8 is the figure of the Rect's posture information and the Scissor's posture information parsed into XML document in accordance with the MPEG-U part2 standard. In order to verify the validity of the parsed XML document, the verification was done using

the MPEG-U part2 reference software. First, convey the parsed XML document to the MPEG-U part2 reference software as in the figure 9-(a) and 10-(a). After that, the corresponding coordinate values are detected as in the figure 9-(b) and 10-(b) through unmarshal task and print that detected coordinate values to GUI Of the user interaction interface system as in the figure 9-(c) and 10-(c). Like this, the XML document in accordance with the MPEG-U standard can be conveyed to a different device and that different machine can obtain the same result if it is equipped with the XML parser which suits the standard.
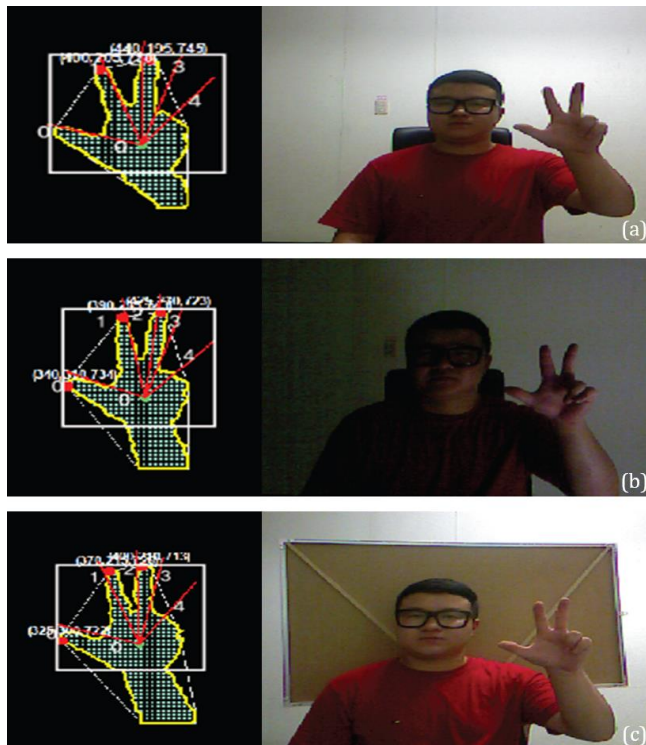

Figure 6. (a) Recognition result of the hand posture under the bright background, (b) recognition result of the hand posture under the dark background, (c) recognition result of the hand posture under the background having the similar color with the human skin
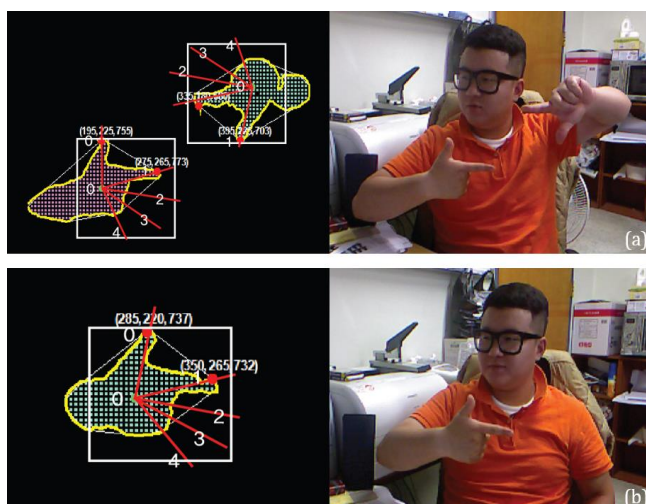

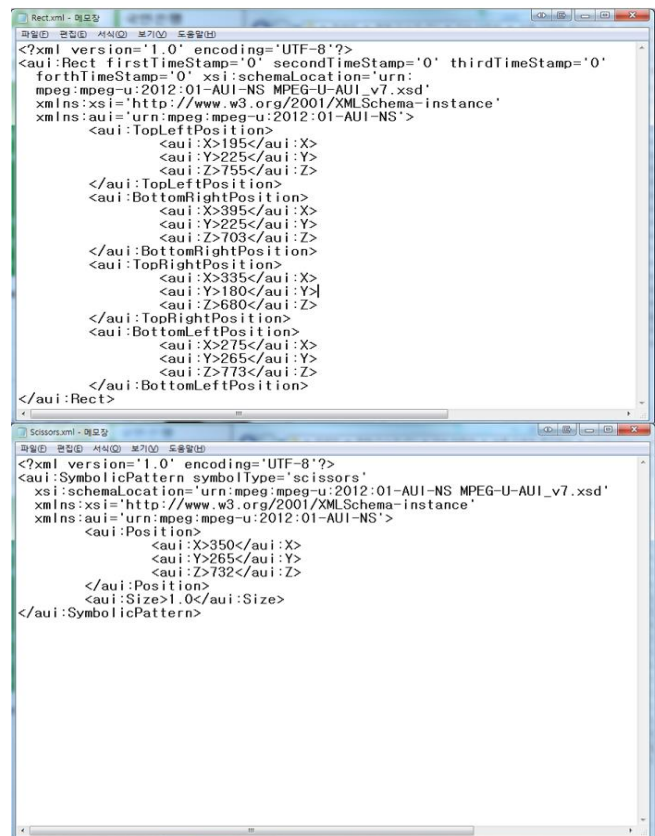**Figure 7.** (a) Hand posture of Rect, (b) Hand posture of Scissors

Figure 8. XML document in accordance with the MPEG-U part 2 standard (a) Rect, (b) Scissors
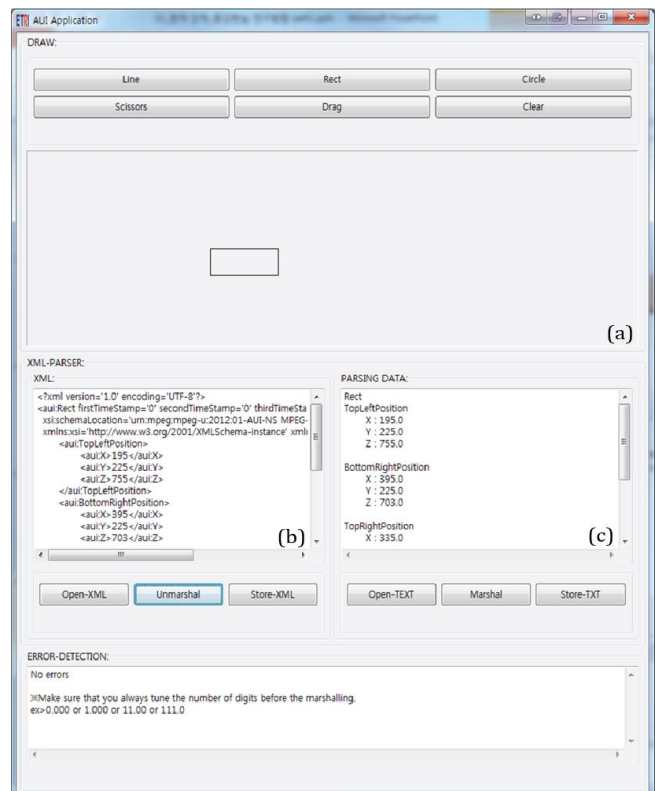

Figure 9. X-Y coordinate output and GUI output from the XML document of Rect
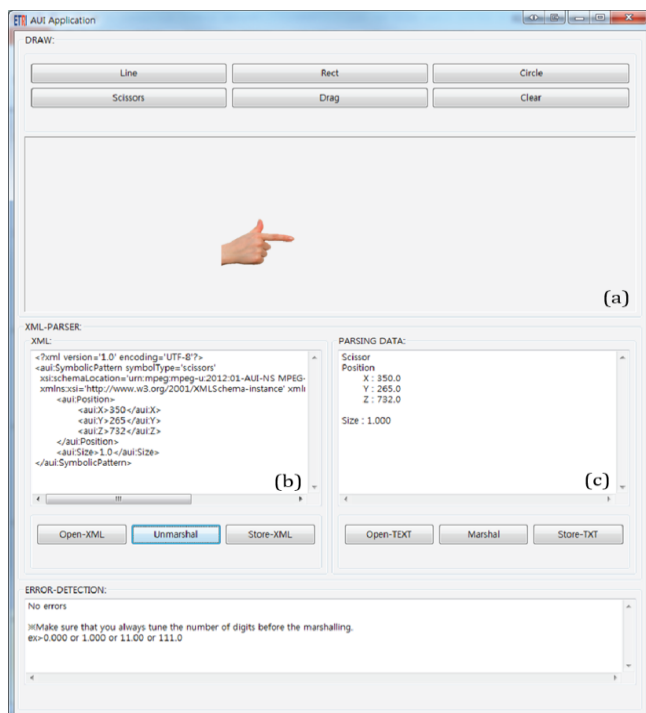
Figure 10. X-Y coordinate output and GUI output
from the XML document of Scissors

IV. CONCLUSIONS

The system proposed in this thesis is the MPEG-U based advanced user interface system which works by way of the hand posture recognition using the depth camera. The hands and the fingers were more robustly recognized using the depth camera, the proposed algorithm to classify the fingers could classify robustly in the change of the various movement and rotations and it could also classify robustly in the folding and unfolding actions. Also, the different scene techniques were able to be interoperable as the standard basis based on the MPEG-U. There also showed standard based interface between the user interface device and the scene technique device. The proposed system will be able to be utilized as the modality in the multi-modal interface system afterwards. Also, if the development of technology utilizing MPEG-U is promoted for the smart TV and Smart phone, more various interfaces will be realized.

ACKNOWLEDGMENT

REFERENCES

[1] Seok-Ju Hong and Chil-Woo Lee, "Human-Computer Interaction Survey for Intelligent Robot," The Korea Contents Society, Vol. 4, No. 2, pp. 507-511, Feb. 2006.

[2] Anastasios Roussos, Stavros Theodorakis, Vassilis Pitsikalis and Petros Maragos, "Hand tracking and affine shape-appearance handshape subunits in continuous sign language recognition," ECCV Workshop on Sign, Gesture and Activity, Hersonissos, Crete, Greece, Sep. 2010.

[3] Yuh-Rau Wang, Wei-Hung Lin, and Ling Yang, "A novel real time hand detection based on skin-color," Consumer Electronics (ISCE), IEEE 17th International Symposium on, pp. 141~142, Jun. 2013.

[4] Xintao Li, Can Tang, Chun Gong, Sheng Cheng and Jianwei Zhang, "Hand Segmentation Based on Skin Tone and Motion Detection with Complex Backgrounds," Chinese Intelligent Automation Conference, Springer Berlin Heidelberg, Vol. 256, pp. 105~111, Jan. 2013.

[5] Robert Y. Wang and Jovan Popovi', "Real-Time Hand-Tracking with a Color Glove," ACM Transactions on Graphics, Vol. 28, Issue. 3, No. 63, Aug. 2009.

[6] R. Lockton and A. Fitzgibbon, "Real-time gesture recognition using deterministic boosting," BMVC, pp. 1~10, Sep. 2002.

[7] V. Argyros and S. Sclaroff, "Database indexing methods for 3D hand pose estimation," Gesture Workshop, pp. 288~299. Apr. 2003.

[8] Candescent NUI Samples & Source code http://candescentnui.codeplex.com/SourceControl/latest#CCT.NUI.Visual/ClusterLayer.cs.

[9] S. Malik, "Real-time hand tracking and finger tracking for interaction," CSC2503F Project Report, Department of Computer Science, University of Toronto, Dec. 2003.

[10] C. Davatzikos and J. L. Prince, "Convexity analysis of active contour problems," Image Vision Computing, Vol. 17, pp. 27~36, Jan. 1999.

[11] I. Oikonomidis, N. Kyriazis and AA. Argyros, "Efficient Model-based 3D Tracking of Hand Articulations using Kinect," BMVC, pp. 101.1~101.11, Sep. 2011.

[12] Junyeong Choi, Hanhoon Park and Jong-ll Park, "Hand shape recognition using distance transform and shape decomposition," Image Processing(ICIP), pp. 3605~3608, Sept. 2011.

[13] Information technology - Rich media user interfaces - Part 2: Advanced user interaction (AUI) interfaces. - ISO/IEC 23007, Feb. 2012.

[14] Gukhee Hand, A-Ram Baek, Haechul Choi, "MPEG-U part2 based advanced user interaction interface system," The Korea Contents Association Journal, Vol. 12, No. 12, pp. 54~62, Dec. 2012.

[15] KINECT. http://en.wikipedia.org/wiki/Kinect.

[16] Sara Taskinen and David I, "Robust estimation and inference for bivariate line‐fitting in allometry," Biometrical Journal, pp. 652-672, Jun. 2011.

[17] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Distance Transforms of Sampled Functions," Theory of Computing, Vol. 8, pp. 415-428, Sep. 2012.

Gukhee Han
March 2006 ~ February 2012 : Graduated from the department of multimedia engineering, Hanbat National University (Bachelor of engineering). March 2012~ Present : Master's degree courses at professional Information and communication school of graduates, Hanbat National University.

Haechul Choi
March 1993~ February 1997 : Graduated from electronic engineering department of Kyeongbuk university (Bachelor of engineering). March 1997~ February 1999 : Graduated from the electricity and electronic engineering department of KAIST (Master of engineering) March 1999~ August 2004 : Graduated from the electricity and electronic engineering department of KAIST. (Doctor of engineering) September 2004~ February 2010 : Senior researcher in the broadcasting media research department of Korea Electronic communication research center. March, 2010~ Present : Assistant professor at the information communication engineering department of Hanbat university.