A Method to Specify a Region of Character String in Augmented Reality

Takashi Higasa*, Shin-ichi Ito*, Momoyo Ito⁺, Minoru Fukumi⁺

Tokushima University
2-1, Mimami-Josanjima, Tokushima, 770-8506, Japan

*{c501737039, s.ito}@tokushima-u.ac.jp
+{momoito, fukumi}@is.tokushima-u.ac.jp

Abstract

This paper proposes a method to input characters and/or character string in an augmented reality using gesture motion. The proposed method detects the region of character string using gesture motion. It consists of five phases; template generation, skin detection, hand region detection, gesture motion extraction and designation of character string region. The template image consists of two fingers because a gesture is to take hold the tips of the first and second fingers. In the skin detection, we extract the skin color on the basis of values in saturation by using threshold processing. The hand region is detected by calculating areas and detecting the area with the maximum value as a hand. The gesture motion is extracted using template matching. In order to show the effectiveness of the proposed method, we conduct experiments.

Keywords: Augmented Reality, HSV Color System, Gesture Motion. Quadrangle for character string region.

1. Introduction

Studies related to virtual reality, augmented reality (AR) and mixed reality have received a lot of attention in recent years. AR is a technology to expand reality environment by adding information. AR technology is aimed at constructing a new user interface and making it useful for work support and information posting. Furthermore, AR service is expected to be applied to various fields such as medical, education, construction, tourism and entertainment. Although head mount displays (HMDs) for AR lack image quality and clarity compared with display monitors, they can add information from a distant position to a wall surface, a floor surface and so on in real environment. Then, these interfaces can be utilized as a GUI (Graphical User Interface). These operations are intuitive. In AR contents, manipulation

of virtual objects often uses hands. The manipulation performed by human hands is intuitive. Then, users often use the soft keyboard in AR when searching information that is key words, characters, words, character string, and so on. If characters and/or character string are retrieved by specifying the region of character string using gesture motion, the issue of the soft keyboard is solved. The final goal of our studies is to create a character string retrieval system. In the proposed system, the region of character string is specified by gesture motion. It is important to extract a gesture motion to construct the proposed system. There are lots of techniques to extract the gesture motion⁽¹⁾. In general, the techniques for gesture motion extraction include skin color detection⁽¹⁾, hand detection^(2,3), fingerprint detection⁽⁴⁾, noise reduction, gesture recognition, fingerprint tracking⁽⁵⁻⁷⁾, and so on. The proposed method to specify the region using a gesture motion consists of five phases; template generation, skin detection, hand region detection, gesture motion extraction and designation of character string region. In order to show the effectiveness of the proposed method, we conduct experiments.

2. Proposed Methods

We propose a method to specify a range of character string by a gesture motion to perform a string search in AR.

Fig. 1. Gesture motion to specify the region of character string. (a) and (b) show the region designation and the gesture motion, respectively.

Fig. 2. Flowchart of the proposed method.

The proposed method sets up two points by the gesture motion. The gesture motion is to take hold the tips of the first and second fingers shown in Fig.1. Then, the region of quadrangle that the two points are vertex of quadrangle is specified. The proposed method detects the region of character string using a gesture motion. It consists of five phases; template generation, skin detection, hand region detection, gesture motion extraction and designation of character string region. The flowchart of the proposed method is shown in Fig. 2.

2.1. Template Generation

The proposed method extracts the gesture motion using a template matching method. The template is created by taking a picture. Then, the first and second fingers are taken hold. The background color is white when taking a picture. The size of the template image is 130×60 pixels.

2.2. Skin Color Detection

The skin color is extracted to recognize hands. The skin color area is extracted using the HSV color system and threshold processing, because the skin color is mixed color. The proposed method employs the HSV color system to detect the skin color. The HSV color system is composed of three components of hue, saturation and brightness. Fig. 3 shows a sample image of HSV color system conversion. In this paper, the skin color is extracted by using values in saturation. The threshold of the saturation for extracting the skin color is $S_{min} < S < S_{max}$. S_{min} and S_{max} are determined according to development environment.

Fig. 3. HSV color system conversion. (a)-(d) are an input image, a hue image, a saturation image and a value image, respectively.

Fig. 4. Contour extraction. The red circle is a starting point.

2.3. Hand Region Detection

The proposed method carries out contour extraction to recognize hands (shown in Fig.4). Firstly, the saturation image is converted into a binary image. The object is represented with white and the background is represented with black. Secondly, the binary image is scanned sequentially from the upper left. Then when finding a white pixel, the pixel is taken as the first outline pixel and the starting point to detect outline. Third, the region contiguous to the starting point is searched to find outline pixels counterclockwise. Then, the first object area is set as the first outline pixel. Finally, the closed curve obtained is taken as the contour when it returns to the starting point.

The extracted contours are stored in a hierarchical structure (shown in Fig 5). (a)-(c) are an input image, a

(a) input (b) Result example

hierarchy	Contour structure
1	1 $\xrightarrow{h_next}$ 2
2	3 $\xrightarrow{v_next}$ 4
3	5 $\xrightarrow{h_next}$ 6
4	7 $\xrightarrow{v_next}$ 8

(c) Hierarchical structure

Fig. 5. Contour detection algorithm. (a)-(c) are an input image, a resulting example of extracted the contours and a hierarchical structure in the contour extraction method.

resulting example and a hierarchical structure, respectively. First, the outermost contour is extracted from the image. The extracted outermost contours are stored in level 1 of the hierarchical structure. Next, the outline in the outermost contour is extracted. If there is a contour, it is stored in level 2 of the hierarchical structure and this operation repeated. In this paper, we refer to the information of the hierarchy of Level 1 because the outermost contour shows a hand. The each contour-enclosed superficial content in the hierarchy of Level 1 is calculated. We regard the region of the maximum superficial content as a hand region.

2.4. Gesture Motion Extraction

Gesture motion is detected by using a template matching method. The proposed method has pre-processing for easy template matching. The pre-processing consists of two phases; blurring processing and processing of designating region to be searched by template matching. The Gaussian filter is used to make blurry image. Using the Gaussian filter, a filtering value of luminance values is calculated on around the target pixel. In other words, we

Fig. 6. Hand region detection.

calculate the luminance value of the target pixel using a function of Gaussian distribution. The function of the Gaussian distribution is expressed by equation (1).

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \quad (1)$$

where x and y are distances from the origin in the horizontal axis, and the vertical axis, respectively. σ is the standard deviation of the Gaussian distribution.

The proposed method narrows the searching region for template matching. Then, the centroid of the detected hand region is calculated. Moreover, the radius R from the centroid is calculated (shown in Fig.6.) as follows;

$$R = \alpha \sqrt{\frac{S}{\pi}} \quad (2)$$

where S and α are the hand region and a bias to create the searching region, respectively. The proposed method is directed to the quadrangle surrounding the circle in which the radius is R (shown in Fig.6.).

The template matching is used to detect the gesture motion. In this paper, Normalized Cross-Correlation (NCC) is used for calculation of similarity. The gesture motion is extracted when the similarity is larger than a certain value. The function of the NCC is expressed by equation (3).

$$R_{NCC} = \frac{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} I(i, j) T(i, j)}{\sqrt{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} I(i, j)^2 \times \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} T(i, j)^2}} \quad (3)$$

where $T(i, j)$ and $I(i, j)$ are the luminance values of the template image and a frame image. M and N are the width and the height of the frame image, respectively.

2.5. Determination of character string region

The start and end points to specify a character string region are determined using a gesture motion. The algorithm for determining the start and the end points is shown in Fig.7. We visualize the start and the end points determined in the

Fig. 7. Procedure of determination of start and end points.

Fig. 8. Determination of a start and an end points. The red and the green circles are the start and the end points, respectively. White quadrangle shows the specified region.

user interface. The start and the end points are displayed in the red and the green circles, respectively (Fig.8). The character string region is the quadrangle surrounded by the start and the end points (shown in Fig.8). Therefore, white quadrangle shows the specified region.

3. Experiments

In order to show the effectiveness of the proposed method, we conducted experiments.

3.1. Experiment Environment

The subjects were 4 healthy subjects (average age: 22 years old). The subject sat on a chair and conducted a task for experiments in our laboratory. The task was to select the

region of printed journal title by a gesture motion. The number of the experiments per him/her was five. The number of iterations per experiment was free. He/she could repeatedly perform to select the region until a successful conclusion. The size of words of the journal title was 300×50 pixels. The line spacing was about 15 pixels. S_{min} and S_{max} for skin detection were 200 and 255, respectively. α for hand region detection was 1.6. σ for the Gaussian filter was 11. Moreover, the threshold of template matching (R_{NCC}) was 0.983.

3.2. Evaluation Method

This paper employed operation accuracy for specifying the region to evaluate the proposed method. The evaluation methods made a comparison between the results of specified region using mouse operation and the proposed method. In operation evaluation, we regarded it as success when the coordinate differences between the mouse operation and the gesture operation on both x -coordinate and y -coordinate became 35 pixels or less.

$$\begin{aligned}
 |M_{Sx} - P_{Sx}| &\leq 35 \\
 |M_{Sy} - P_{Sy}| &\leq 35 \\
 |M_{Ex} - P_{Ex}| &\leq 35 \\
 |M_{Ey} - P_{Ey}| &\leq 35
 \end{aligned} \tag{4}$$

where M and P show the mouse operation and the proposed method, respectively. S , E , x and y are start position, end position, x -coordinate and y -coordinate, respectively.

4. Experimental Results and Discussions

4.1. Experimental Results

Figure 9 shows the results of the mouse operation and the gesture operation for specifying the region of the journal title. (a)-(d) are show the results of the subject A to D, respectively. Red and green circles and white quadrangle are the same as those in Fig.8. Yellow quadrangles are the results by the mouse operation. Table 1 shows experimental results of the operation evaluation. Table 1 (a)-(d) show the results of the subjects A, B, C and D, respectively. "S or F" represents success or failure. The number of the success times for specifying the region was 17 (subjects A, B and D : 4 times, subject C : 5 times). The number of the failed times was three (subjects A, B and D : one time). In the subject D, the coordinate differences ($M_{Ex} - P_{Ex}$, $M_{Ey} - P_{Ey}$) was small compared with one of other subjects. In the subject B, the coordinate differences became smaller by repeated experiments.

Fig. 9. Result of the operation for specifying the region of the journal title. (a)-(d) show the results of the subjects A to D, respectively. Red and green circles and white quadrangles are the same as those in Fig.8. Yellow quadrangle are the results by the mouse operation.

4.2. Discussions

We confirmed that the number of the success times was 17. In the proposed method, the skin color was detected by threshold processing on the saturation, hand region was detected using contour detection algorithm and the gesture motion was extracted by template matching. These results suggest that the accuracy of each approach is high, and the parameters for each approach are set appropriately. We confirmed that the number of the failed times are three. It is thought that these results are due to the fact where the region of fingers went out of the screen because the character of the journal title was the edge of the screen. Furthermore, the gesture motion was not easy, and therefore the start and end points were not stable.

In the results of the subject D, the coordinate differences were small because the subject D corrected positioning of start and end points in a special way. That special way was to continue to close the two fingers and move the designated point.

In the results of the subject B, the coordinate differences became smaller by repeated experiments. These results suggest that the subject B adjusted gradually the gesture motion.

5. Conclusions

In this paper, we proposed a method to specify a region

of character string by gesture motion. The proposed method consists of five phases; template generation, skin detection, hand region detection, gesture motion extraction and designation of character string region. In the template generation, we created an image when two fingers touch each other as a template image. The skin color was detected by threshold processing on the saturation in the HSV color system. The hand region was detected using a contour detection algorithm. The gesture motion was extracted by using template matching. In order to show the effectiveness the proposed method, we conducted experiments. In the experimental results, we could obtain good results (the number of the success times was 17). However, we had three failed times. It is thought that these results are due to the fact where the region of fingers went out of the screen because the character of the journal title was the edge of the screen. Furthermore, the gesture motion was not easy, and therefore the start and the end points were not stable.

The future work will improve the way of gesture motion because the gesture motion was not easy in this paper.

References

- (1) Manresa, Cristina, et al. : "Hand tracking and gesture recognition for human-computer interaction." ELCVIA Electronic Letters on Computer Vision and Image Analysis, pp. 96-104, 2005.
- (2) Oliveira, V. A., and A. Conci.: "Skin Detection using HSV color space." H. Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi. pp. 1-2, 2009.
- (3) Sharma, Ram Pratap, and Gyanendra K. Verma.: "Human computer interaction using hand gesture." Procedia Computer Science 54, pp. 721-727, 2015
- (4) Marengoni, Mauricio, and Denise Stringhini. : "High level computer vision using opencv." Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 24th SIBGRAPI Conference on. IEEE, 2011.
- (5) Shapiro, L. G. & Stockman, G. C : "Computer Vision", Prentice Hall, pp. 137, pp. 150, 2001.
- (6) Mark S. Nixon and Alberto S. Aguado. "Feature Extraction and Image Processing", Academic Press, pp. 88, 2008.
- (7) Nagarajan, S., T. S. Subashini, and V. Ramalingam : "Vision based real time finger counter for hand gesture recognition." International Journal of Technology, pp.1-5 , 2012.

Table 1. Results of the operation precision evaluation.

(a) Subject A

	S or F	M_{Sx}	M_{Sy}	M_{Ex}	M_{Ey}	P_{Sx}	P_{Sy}	P_{Ex}	P_{Ey}	$M_{Sx} - P_{Sx}$	$M_{Sx} - P_{Sy}$	$M_{Ex} - M_{Ex}$	$M_{Ey} - M_{Ey}$
1	O	134	161	461	225	116	154	443	190	-18	-7	-18	-35
2	x												
3	O	180	204	499	253	151	175	496	245	-29	-29	-3	-8
4	O	136	194	456	246	132	179	451	236	-4	-15	-5	-10
5	O	136	198	464	245	135	175	457	233	-1	-23	-7	-12

(b) Subject B

	S or F	M_{Sx}	M_{Sy}	M_{Ex}	M_{Ey}	P_{Sx}	P_{Sy}	P_{Ex}	P_{Ey}	$M_{Sx} - P_{Sx}$	$M_{Sx} - P_{Sy}$	$M_{Ex} - M_{Ex}$	$M_{Ey} - M_{Ey}$
1	x												
2	O	135	148	467	205	118	135	435	174	-17	-13	-32	-31
3	O	115	167	436	219	105	148	417	189	-10	-19	-19	-30
4	O	127	178	451	216	106	158	430	206	-21	-20	-21	-10
5	O	124	168	461	222	110	135	451	220	-14	-33	-10	-2

(c) Subject C

	S or F	M_{Sx}	M_{Sy}	M_{Ex}	M_{Ey}	P_{Sx}	P_{Sy}	P_{Ex}	P_{Ey}	$M_{Sx} - P_{Sx}$	$M_{Sx} - P_{Sy}$	$M_{Ex} - M_{Ex}$	$M_{Ey} - M_{Ey}$
1	O	163	172	482	221	154	157	475	194	-9	-15	-7	-27
2	O	165	172	481	222	168	160	457	203	3	-12	-24	-19
3	O	182	165	489	217	185	156	467	207	3	-9	-22	-10
4	O	156	155	484	203	147	136	461	196	-9	-19	-23	-7
5	O	159	155	478	196	147	152	445	208	-12	-3	-33	12

(d) Subject D

	S or F	M_{Sx}	M_{Sy}	M_{Ex}	M_{Ey}	P_{Sx}	P_{Sy}	P_{Ex}	P_{Ey}	$M_{Sx} - P_{Sx}$	$M_{Sx} - P_{Sy}$	$M_{Ex} - M_{Ex}$	$M_{Ey} - M_{Ey}$
1	O	158	164	482	214	159	165	470	213	1	1	-12	-1
2	O	144	175	472	232	141	177	464	232	-3	2	-8	0
3	O	135	190	458	237	133	188	443	232	-2	-2	-15	-5
4	O	148	200	467	250	153	202	470	258	5	2	3	8
5	x												