Hand Gesture Recognition Algorithm using Mathematical Morphology

Jong Ho Choi[1] and Duck Young Ko[2]

[1]Department of Electronic Engineering, Kangnam University,
Kiheung, Yongin, Kyounggi, 449-702, Korea
Tel. 82-031-280-3530, Fax.: 82-031-280-3535
e-mail: jhchoi@kangnam.ac.kr
[2] Department of Electronic Engineering, Jeon-Ju Technical College,
Hyoja, Wansan, Jeonju, Jeonbuk, 560-760, Korea
Tel. 82-855-2884, Fax.: 82-063-220-3709
e-mail: dyko@jtc.ac.kr

Abstract: In this paper, we present a fast algorithm for hand gesture recognition of a human from an image by using the directivity information of the fingers. To implement a fast recognition system, we applied the morphological shape decomposition. A proposed gesture recognition algorithm has been tested on the 300 x 256 digital images. Our experiments using image acquired image camera have shown that the proposed hand gesture recognition algorithm is effective.

1. Introduction

The use of gestures provides an attractive alternate to cumbersome interface devices for human-computer interaction. This has motivated a very active research area concerned with computer vision-based analysis and interpretation of hand gestures[1]. The most important issues in gesture recognition are the simplification of algorithm and the reduction of processing time. The mathematical morphology based on geometrical set theory is best used to perform the processing. A key idea of proposed algorithm in this paper is to apply morphological shape decomposition. The primitive elements extracted from a hand gesture include in very important information on the directivity of the hand gestures.

Based on this characteristic, we proposed the morphological gesture recognition algorithm using feature vectors calculated to lines connecting the center points of a main-primitive element and sub-primitive elements. We are also proposed the gesture recognition algorithm using chain coding of a trajectory obtained from image gesture sequences.

In order to evaluate the accuracy of the proposed gesture recognition process, the experiment using actual image sequence is performed. Coupling natural interactions such as hand gesture with an appropriately designed interface is a valuable and powerful component in the building of TV switch navigating and video contents browsing system. Research efforts will focus on further improvements of the interface to work with gestures in a real system implementation[2].

2. Preprocessing

The morphological image segmentation algorithm is based on the dynamic thresholding and background subtraction. The basic principle is to eliminate the lighter or darker background and the smaller and larger size particles. The procedures are performed as adopting a top-hat transform and opening operation

The opening operation makes the edges of object soft and removes objects which are smaller than the structuring elements, just like a filter. The closing operation can fill out the small holes which is represent on the valley style object. Thru this theory we can find that the opening and closing operations are used like a low pass filter to remove positive and negative noise components. During this study we used what has been defined as a low pass filter in this manner[3].

$$f_{lp}(m,n)=f(m,n)\bigcirc B_1(m,n)\bullet B_1(m,n)\bigcirc B_2(m,n)\bullet B_2(m,n)\ldots\ldots \quad (1)$$

Where f is the binary image and B is the structuring element.

The two-level thresholding of image is defined as following form.

$$f_t(i,j)= \begin{cases} 1 & \text{for } f(i,j) \geq t \\ 0 & \text{for } f(i,j) < t \end{cases} \quad (2)$$

Where t is a threshold value.

3. Shape Decomposition

The shape decomposition is to find a set of open sets $\{X_1, \ldots\ldots, X_n\}$ whose union is X:

$$X= \bigcup_{i=1}^{n} X_i, \qquad X_i \in G(R^2) \quad (3)$$

Where $G(R^n)$ is the class of the topological open sets on the Euclidean space R^n and $X \in G$ is an open nonempty set containing no half space.

The whole procedure of morphological shape ompositon is given by the following recursive relation[4,5,6].

$$\begin{aligned} X_i &= (X-X'_{i-1})n_iB \\ &= ((X- X'_{i-1}) \ominus n_iB) \oplus n_iB \\ X_i &= \cup X_j \qquad (1 \leq j \leq i) \\ X_0 &= \emptyset \end{aligned} \quad (4)$$

Stopping Condition $(X-X'_k) \ominus B)= \emptyset$

Where r_i is the radius of the maximal inscribable objects r_iB and B is a structuring element.

In the conventional algorithms, a structuring element is only used[4]. However, if we just use a single element, the descriptive efficiency and intuitive recognizability are decreased. Because of these limitations, we proposed a shape decomposition algorithm using the rohmbus, circle, square as structuring elements.

The method of using the 4 structuring elements can be improved coding efficiency, however, it has a critical weakness because of incresing of processing time. To solve the this problem, we proposed the method obtaining a structuring element that is similar to shape characteristics of object. The resulting description efficiency of a morphological shape decomposition is highly dependent on the preselection of a structuring elements.

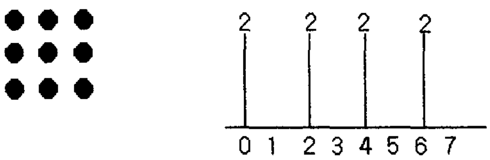
The chain coding of an object itself is a good shape descriptor, but it does not provide shape information of an object. Because of this consideration, the chain code is applied to the shape matching method between the object and sturcturing elements. Fig. 1 shows the 8-directional chain code and shape function of structuring elements. Base on this shape function analysis, a structuring element is preselected through matching the highest similarity of shape function between structuring elements and object.

3	2	1
4	(i,j)	0
5	6	7

(a) 8-directional chain code

(b) rohmbus shape function

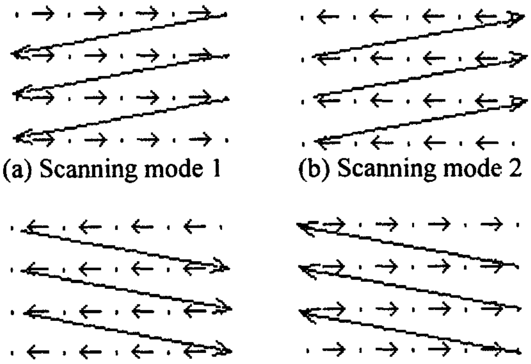
(c) 5 × 5 circle shape function



(d) 3× 3 square shape function

Fig. 1. Chain code and shape function of structuring elements

We have discussed the algorithm for shape ecomposition of a grayscale image using multiple structuring elements. On the contrary, using classical scanning order, from left to right and top to bottom, when structuring a compact object, generally leads to residues located at the right and the bottom of object boundaries. To eliminate this problem, we applied a solution which extends the operation to four scanning orders and it used in a algorithm for shape decomposition. To perform partial overlapping erosion efficiently four kinds of scanning are used as shown in Fig. 2.



(a) Scanning mode 1 (b) Scanning mode 2 (c) Scanning mode 3 (d) Scanning mode 4

Fig. 2. 4 scanning modes

The shape decomposition procedure progressively extracts the description points and operates in the following way. Firstly, the preselected structuring element is translated point by point over the binary image according the first scanning mode. For each position of structuring element, a largest structuring element that belongs to object is extracted. Then the translation of a structuring element goes on following the same scanning mode. The same procedure is reiterated for the four scanning modes.

4. Gesture Recognition

4.1 Gesture Recognition using Geometric Feature Vector

Primitive elements decomposed by morpholo-gical shape decomposition are extracted in order of size and represented hierarchically by their adjacent relation described as follows:

step.1 Define the largest primitive elements X_i as a main element.

step.2 Sort sub-elements X_i $(i=2,3,4,...n)$ in order of size

step.3 Eliminate elements which are not appropriate to recognize the hand gestures. (e.g. As depicted in Fig. 3, $X_i(i=3,4,...n)$ is eliminated if the distance d_i between center points of X_i and X_1 is not larger than d_{i-1}.)

step.4 Repeat step.3 from $i = 2$ to n, where n is the number of times to repeat the shape decomposition until recognizable feature vectors of a hand-gesture are extracted.

An algorithm proposed in this paper is to recognize hand gestures by using feature vectors representing the relation of the decomposed elements. The feature vectors represent a set of lines which connect the center points of a main-primitive element to sub-primitive elements. The feature vectors is defined as angles of its lines

$$\mathbf{x} = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_{n-1}\} \quad (5)$$

where n represents the number of primitive elements. And θ is defined as

$$\theta_{n-1} = \tan^{-1}\left(|y - y_{n-1}| / |x - x_{n-1}|\right) \quad (6)$$

where (x, y) represents the center point of its primitive elements. The average of the feature vectors is defined as

$$\bar{\mathbf{x}} = (\theta_1 + \theta_2 + \theta_3 + \Lambda + \theta_{n-1})/(n-1). \quad (7)$$

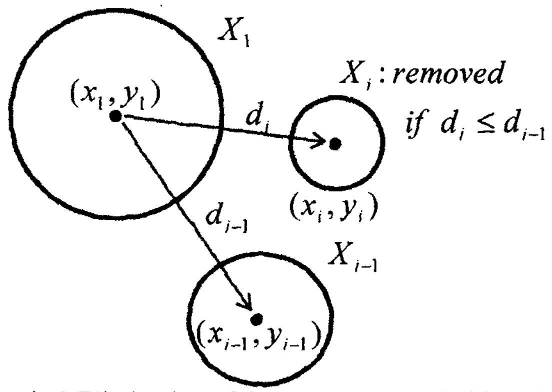


Fig.3 Elimination of the unnecessary primitive element

The feature vector derived from (6), (7) is the information about the directivity of the finger. Therefore, a hand-gesture pointing region#1 is recognized in case feature vector is in 45°~134°, region#2 is in 135°~225°, region#3 is in 224°~43° respectively in Fig.4. An hand-gesture region#4 is the case the sub primitive elements are located near the main primitive element. By thresholding the distance between them, sub-primitive elements are removed and this case is recognized as region#4.

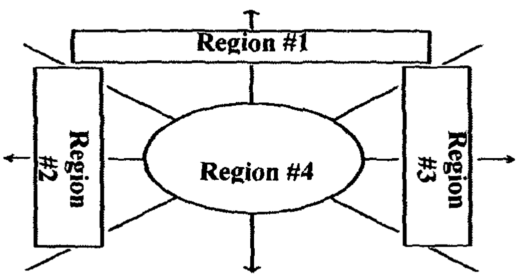


Fig.4 Recognition of the feature vector

To detect a hand region, the RGB image taken by camera is converted to YC_bC_r space and thresholded it by the skin color range ($77 \le C_b \le 127$, $133 \le C_r \le 173$) and the noise and holes are filtered by morphological operation in Fig.5.

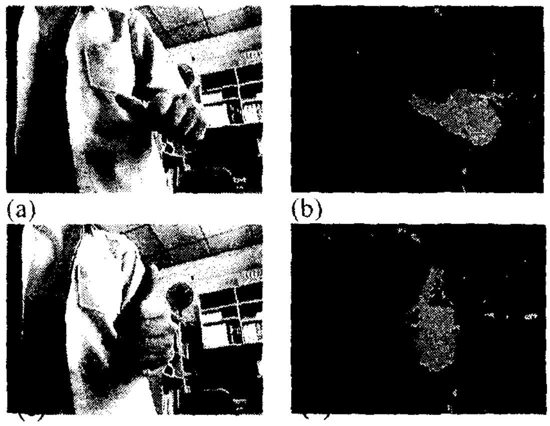


Fig.5 source images (a), (b) and skin region detected images (c), (d).

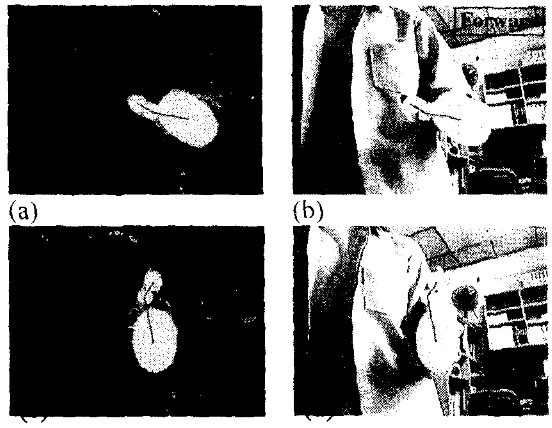


Fig.6 (a), (c) are shape decomposed image and (b), (d) are recognized results by feature vectors.

Fig.6 shows that (a), (c) are decomposed and (b), (d) are recognized forward, play function respectively by vectors implicating the finger's direction. <Table 1> is mean values of the hand-gestures estimated.

<Table 1> Mean angles of the feature vectors

Hand Gesture	Function	Mean angle
Region#1	Play	93.2°
Region#2	Forward	172.8°
Region#3	Rewind	23.4°
Region#4	Stop	x

4.2 Gesture Recognition using Geometric Trajectory

The gesture is recognized by using the feature vectors characterized as chain code edge vectors that consist of trajectory of center points in primitive elements extracted by shape decomposition. The key idea of proposed algorithm is to apply the orientation-adaptive image processing. The directivity of trajectory elements can be extracted by chain coding. This feature vectors are used in the phase of gesture recognition.

In this case, the center points of primitive elements in a gesture image sequence are extracted in the first place. In the second step, we acquire a trajectory connecting the center points. In the next step, we pick one starting point among the center points of primitive elements and assign an angle to the segements connecting every pair of end points in clockwise direction. In this case, the adjacent direction angles appear to be multiples of 45° and the angles can be encoded by a number in set {0,1,...,7}. Therefore, we can simply consider the chain code distribution of a trajectory instead of the real angle distribution. These eight variables compose the chain code edge vectors as $x=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$. A chain code edge vector is normalized as follows and the normalized edge vector will be used for gesture recognition.

$$X = x / (x_1^2+x_2^2+x_3^2+x_4^2+x_5^2+x_6^2+x_7^2+x_8^2)^{1/2} \quad (8)$$

The chain code edge vectors provide the information on shape orientation. From the vectors, the major directivity of shape can be determined. In the last step, a gesture is recognized by using the ratio of chain edge vectors.

A proposed gesture recognition algorithm has been tested on the two 300 x 256 digital images sequence. The binary image sequences obtained after performing the operations of background subtraction, noise removal, and thresholding, respectively, are shown in Fig. 7.

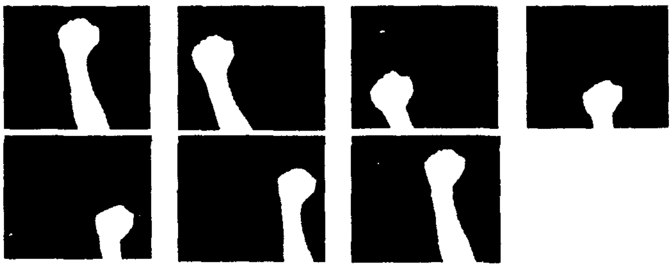


Fig. 7. Gesture image sequences used in experiment

The hand gesture sequences are decomposed to disks by using a 5x5 circle structuring element. From the results we chosed the largest primitive elements. To detect the center points, the erosion operation is performed respectively until To recognize the gesture from an image sequence, the chain edge vectors are calculated by using the chain code obtained from the trajectory of center points. The results are shown in <Table 2>.

<Table 2> Value of chain code edge vectors for each geesture

Gesture	Elements of chain code edge vector							
	X_1	x_2	x_3	x_4	X_5	x_6	x_7	x_8
Circle	107	32	69	45	73	51	72	34
Triangle	164	0	46	89	2	75	56	3
Gesture	Element of normalized chain code edge vector							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Circle	0.58	0.18	0.38	0.25	0.40	0.28	0.40	0.19
Triangle	0.77	0.00	0.22	0.42	0.01	0.35	0.26	0.01

To recognize the gestures from above results, we compare the ratio of edge vectors to its of ideal gesture. The normalized chain code edge vectors calculated from ideal gestures are shown in <Table 3>. Based on the edge vector analysis, the gesture can be determined through matching the highest similarity of edge vectors between the ideal gesture and the measured gesture. We can be recognize correctly the gestures from the results.

<Table 3> Value of chain code edge vector for each ideal esture

Gesture	Elements of chain code edge vector							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Circle	68	34	55	36	53	36	55	32
Triangle	160	0	54	80	0	80	54	0
Gesture	Element of normalized chain code edge vector							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Circle	0.52	0.26	0.42	0.28	0.41	0.28	0.42	0.25
Triangle	0.76	0	0.26	0.38	0	0.38	0.26	0

5. Conclusions

In this paper, we proposed the morphological hand-gesture recognition algorithm that uses the relation of the locations of primitive elements extracted from a hand gesture. A key idea of this algorithm is to recognize by using feature vectors calculated with lines connecting the center points of a main primitive element and sub primitive elements.

Through experiments, we showed that four hand gestures for video contents navigation and image sequences are recognized exactly using morphological shape decomposition. Therefore, this algorithm can be applied various navigation of video data and other interfaces of the control systems. However, It is difficult to adopt this to more complex interface. For more various applications, a research about more various hand gesture recognition should go on.

References

[1] J. Yamato, J. Ohya and K. Ishii, "Recognition human action in time-sequential images using hiddenMarkov model", Proc. CVPR'92, 1992.

[2] T. Starner, J. Weaver and A.Pentland, "Real time American Sign Language recognition using desk and wearable computer based video", IEEE Trans. PAMMI, Vol.20, No.12, 1998.

[3] Skolnick, M.M., "Application of Morphological Transformations to the Analysis of Two-Dimensional Electrophoretic Gels of Biological Materials", Computer Vision, Graphics and Image Processing, Vol.35, 1986.

[4] Pitas, I. and Venetsanopoulos, A. N., "Morphological Shape Representation", Pattern Recognition, Vol.25, No. 6, 1992.

[5] Pitas, I. and Venetsanopoulos, A. N., "Morphological Shape Decomposition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12, No.1, 1990.

[6] Serra, J., "Introduction to Mathematical Morphology", Computer Vision, Graphics, and Image Processing, Vol.35, No.3, 1986.