International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.9 (2015), pp.281-290
http://dx.doi.org/10.14257/ijmue.2015.10.9.29

A Robust Hand Gesture Recognition System Using Hierarchical Features Based on Infrared Image

Hye-Youn Lim[1], Jung-Hyun Park and Dae-Seong Kang[1]

[1]*Department of Electronics Engineering, Dong-A University, 37 Nakdong-Daero 550, beon-gil saha-gu, Busan, Korea*
dskang@dau.ac.kr

Abstract

A hand gesture recognition system for American sign language (ASL) using hierarchical features based on an infrared image is proposed. To reduce the error rate and illumination chage, the infrared image is used in this article. Hierarchical features consist of object extern, Hu-moment invariants, and direction features. First, circularity and eccentricity can be computed from the object extern feature. And then, ASL is classified by K-means using them. Next, the moment invariants features are used to recognize hand gestures by back-propagation (BP). Finally, the direction feature can accurately classify similar gestures like G and Z, I and J, U and H. The goal of this article is to achieve an efficient and effective hand gesture recognition system that meets the high recognition rate of gestures. Through experiments, the recognition rate for the proposed method is 97.15% and it takes 0.046 s to process one frame.

Keywords: *ASL, BP, Hand gesture recognition, Hierarchical features, Infrared image*

1. Introduction

A gesture is simple and effective means of communication in human machine interaction (HMI). Recently, living in the development of information technology (IT) the number of people using smart devices is increasing rapidly. Therefore, work on gesture recognition studies actively continues. The gesture can transmit various meanings through hands, face, legs or body and so on [1-2]. In this article, we study the hand gesture which is widely used among gestures.

The hand gesture recognition methods can be divided into two groups according to the contact condition: contact-based method and contactless-based method. Typical contact-based methods are finger sensors method and an electromyogram (EMG) sensor method. The finger sensors method which has sensors each finger extracts the information of moving finger regardless position and direction of hand. The EMG sensor method measures muscular motion. It should exactly recognize hand gestures. However, it is expensive. Also, it is inconvenient and fatigue increases on wearing sensors for a long time [3-6]. The contactless-based method is used for image of hand gesture such as cameras. It is convenient and has low prices. Therefore, the contactless method is widely used in hand gesture recognition. However, it is vulnerable to sudden illumination changes and complex backgrounds.

The hand gesture recognition system proposed in this article employs the contactless method based on camera. To offset weakness in the contactless method, it presents hierarchical features in infrared image based on the Kinect camera for the high recognition [7].

The rest of this article is organized as follows. In Section 2, relevant theories Hu-moment invariants and back propagation (BP) are introduced. The proposed method is described in Section 3. The experimental results and recognition performances are in Section 4. Finally, conclusions are given in Section 5.

ISSN: 1975-0080 IJMUE

2. Relevant Theories

2.1. Hu-Moment Invariants

A method of moment invariants is another of shape descriptor. The shape descriptor is a series of numeric character set represented as pixel and position values of whole image.

In case that two dimensional image with x and y axis is function f(x, y), geometric moments is defined as follows:

$$m_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x^p y^q f(x,y) \tag{1}$$

Where p and q are zeros or grater valid integers than zero, (p+q) is the order of moment N is pixel size in width and M is pixel size in height.

Central moment for image's center of gravity is

$$\mu_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (x-\overline{x})^p (y-\overline{y})^q f(x,y) \tag{2}$$

Where $\overline{x} = (m_{10}/m_{00})$ and $\overline{y} = (m_{01}/m_{00})$.

An advantage of existing moment invariants is that it is invariable to moving object; however, a disadvantage is that it is sensitive to rotation and size change. Therefore, to offset this weakness, seven moment invariants were introduced by Hu in 1962 [8][9].This is consisted of nonlinear combination of central moment normalization below three orders. Moment normalization can be computed from central moment.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}, \quad r = \frac{p+q}{2} + 1. \tag{3}$$

Finally, Figure 1 shows seven Hu-moment invariants using equation 3. Four moment invariants in red rectangle are only used in this article due to light operation.

$$\boxed{\begin{aligned}
I_1 &= \eta_{20} + \eta_{02} \\
I_2 &= (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2 \\
I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
I_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2
\end{aligned}}$$

$$\begin{aligned}
I_5 &= (\eta_{30} - 3\eta_{12})^2 + (\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
I_6 &= (\eta_{20} - \eta_{02})[(\eta_{03} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]
\end{aligned}$$

Figure 1. Hu-Moment Invariants

2.2. Back-Propagation

Back-propagation (BP) is kind of delta learning multilayer neural networks model. It can be solutions when single layer neural networks can't separate dimensions.

Step 1 : Initialize weights and counter
$w \leftarrow 0$ or small random value
P $\leftarrow$ number of training pattern pairs
$k \leftarrow 1$

Step 2 : Set learning rate $\alpha\,(0 < \alpha \leq 1)$

Step 3 : For each training pattern pair (x, d)
do Step 4–7 until k = P

Step 4 : Compare output

$$y = f(NET) = \begin{cases} +1 & ; \quad NET > T \\ 0 & ; \quad NET = T \\ -1 & ; \quad NET < T \end{cases}$$

Step 5 : Compare output and desired output
if y = d , k $\leftarrow k + 1$ and goto Step 3

Step 6 : Update weights

$$\Delta w \leftarrow w + \alpha\,(d_k - y_k)\,x_k$$
$$w \leftarrow w + \Delta w$$

Step 7 : Increase counter and goto Step 3
$k \leftarrow k + 1$

Step 8 : Test stop condition
If no weights changed in Step 3–7, stop
else, $k \leftarrow 1$ and goto Step 3

Figure 2. The Procedure of Back Propagation Algorithm

Figure 2 shows the process of BP algorithm. BP learning system is proceeded by error back propagating and it is largely made up of three steps. First, applying the learning input pattern to the neural networks and calculate output values and then, find out error values by calculating subtraction of output values and desired values. Lastly, back propagating error values to the output and hidden layers for changing weight values. Learning process can take long time but after learning, It shows fast speed in processing [10-13].

3. The Proposed Method

The proposed recognition system is shown in Figure 3. This has three steps. The first step is an execution of the preprocessing for hand region. The second step involves hierarchical features construction. The last step is the classification of the hand gestures using BP. Finally, the proposed recognition system finds the most accurate hand gestures.

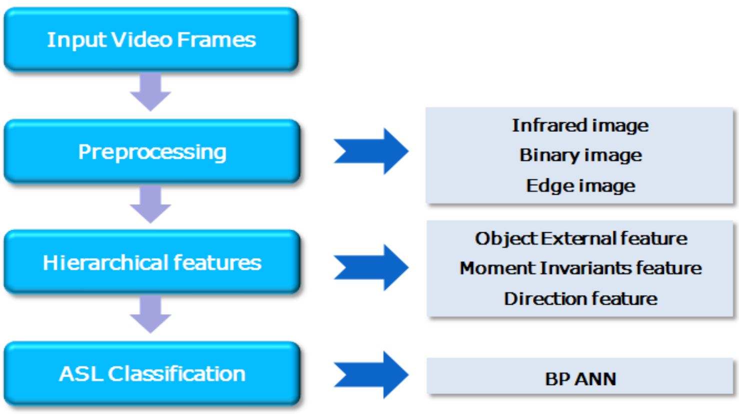


Figure 3. The Process of Proposed Method in this Article

3.1. Preprocessing for Hand Region

A flowchart for preprocessing based on images generation is shown in Figure 4. These images are used in hierarchical features of next steps. RGB image and depth image through infrared transceiver were obtained from kinect. The distance limit is from 55 to 85cm based on camera when you hold up your hand in seating for recognization. Depth image is used for the segmentation, extraction of object(hand), and background area. After binary, detected obejct(hand) is setted to region of interest(ROI) using 4-connected labeling. However, hand region can contain noise types such as shadows and illuminations. Therefore, it needs to remove inner noises usig labeling again. And then, distance transform vector is used here to find the accurate center point in hand. The distance transform vector provides weight as the distance to a specified pixel. As the distance increases, the pixel value increases form 0 to 255. When the pixel value exceeds a the certain level(255), it starts from 0 again [14]. The region under the wrist is removed based on center point. Radius(r) in palm can be computed from center point and weight of brightness. The outside of a green circle with radius ´r´(-25°~-135° region) is removed owing to unnecessary region for gesture recognition. Figure 5 shows the elimination process of the arm region. Figure 6 shows the resulting images of preprocessing.

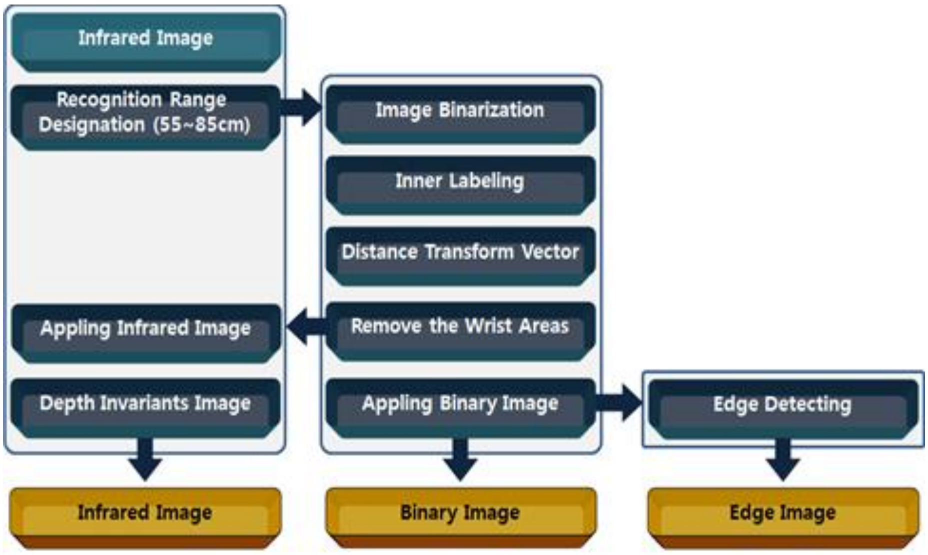
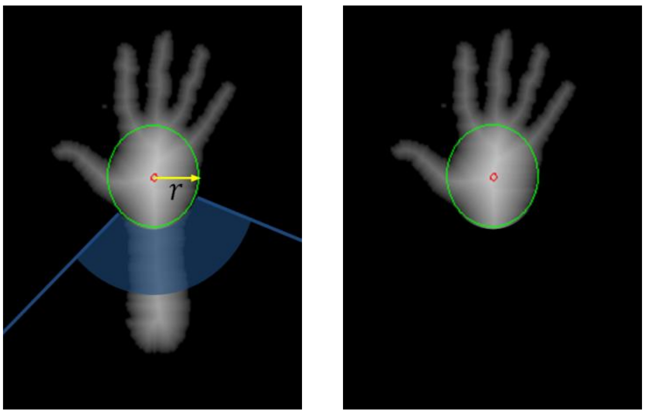


Figure 4. Flowchart for the Preprocessing Based on Images Generation



(a) (b)

Figure 5. The Elimination Process for the Arm Region

(a) Depth (b) Binary (c) Edge

Figure 6. The Resulting Images of the Preprocessing

3.2. Hierarchical Features

Figure 7 shows the flowchart of hand gesture recognition using the hierarchical features. First, circularity and eccentricity can be computed from the object extern feature. And then, ASL is classified by K-means using them. Next, the moment invariants features are used to recognize hand gestures by back-propagation(BP). Finally, the direction feature can accurately classify similar gestures like G and Z, I and J, U and H.

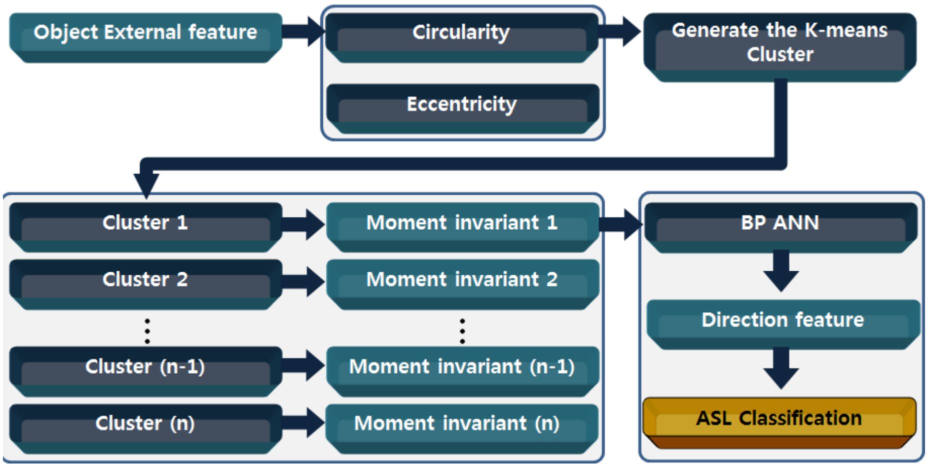


Figure 7. The Flowchart of Hand Gesture Recognition Using the Hierarchical Features

Hierarchical features are computed from obtained images(infrared image, binary image, edge image). Hierarchical features consist of object extern, Hu-moment invariants, and direction feature. First, the object extern feature calculates circularity and eccentricity. Circularity can be defined as

$$\text{Circularity} = (\text{Perimeter})^2/(4\pi \times Area). \tag{4}$$

Where perimeter is the hand external perimeter and Area is hand region.

Eccentricity is a parameter associated with every conic Section. It can be thought of as a measure of how much the conic Section deviates from being circular. The eccentricity can be computed from minor and major axis:

$$\text{Eccentricity} = \frac{\sqrt{a^2 - b^2}}{a} \tag{5}$$

Where a is the length of the semi-major axis and b is the length of the semi-minor axis.

Here, using moment invariant features, the eccentricity can be redefined as follow:

$$\text{Eccentricity} = \frac{(\eta_{20} - \eta_{02})^2 - 4{\eta_{11}}^2}{(\eta_{20} - \eta_{02})^2} \tag{6}$$

Table 1 shows the resulting values of circularity and eccentricity each hand gesture. In case of similar hand gestures, the values of circularity and eccentricity are alike [15-16].

Table 1. The Resulting Values of Circularity and Eccentricity Each Hand Gesture

Hand	Circularity	Eccentricity	Hand	Circularity	Eccentricity
1	0.51	0.52	J	0.5	0.81
2	0.34	0.43	K	0.37	0.41
3	0.26	1.14	L	0.36	1.59
4	0.22	0.91	M	0.67	1.39
5	0.19	1.03	N	0.8	1.33
6	0.26	0.44	O	0.8	0.94
7	0.25	0.7	P	0.46	0.85
8	0.28	0.85	Q	0.37	0.89
9	0.27	0.68	R	0.46	0.4
A	0.86	1.52	S	0.78	1.25
B	0.72	0.49	T	0.74	1.03
C	0.37	0.73	U	0.5	0.28
D	0.52	0.5	V	0.31	0.37
E	0.77	0.75	W	0.28	0.57
F	0.59	0.51	X	0.52	0.8
G	0.56	0.58	Y	0.38	1.33
H	0.54	0.33	Z	0.52	0.51
I	0.53	0.82			

The circularity sets x-axis and the eccentricity sets y-axis. And then ASL is clusterd by K-means. The K-means is an intuitive algorithm based on distance [17]. In this article, the number of clustering sets 10. Figure 8 shows the clustering result of ASL by K-means.

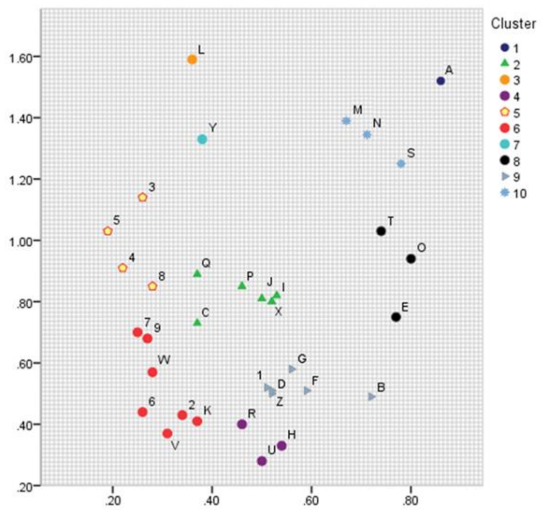


Figure 8. The Clustering Result of the ASL by K-Means

Table 2. The Cfor Correcting the Angle

$\eta_{20}-\eta_{02}$	η_{11}	Θ	Range of θ
zero	zero	0˚	
zero	+	+45˚	
zero	-	-45˚	
+	zero	0˚	
-	zero	-90˚	
+	+	$\theta=\frac{1}{2}arctan\frac{2\eta_{11}}{\eta_{20}-\eta_{02}}$	0˚~45˚
+	-	$\theta=\frac{1}{2}arctan\frac{2\eta_{11}}{\eta_{20}-\eta_{02}}$	-45˚~0˚
-	+	$\theta=\frac{1}{2}arctan\frac{2\eta_{11}}{\eta_{20}-\eta_{02}}+90^\circ$	45˚~90˚
-	-	$\theta=\frac{1}{2}arctan\frac{2\eta_{11}}{\eta_{20}-\eta_{02}}$	-90˚~-45˚

In the clusters included two or more elements, moment values are computed from Hu-moment invariants. ASL is classified according to Momentum BP ANN using I_2, I_3, I_4 among moment values. In case of similar moment values, some gestures among ASL are additionally classified according to the direction feature. The direction feature is used in the major and minor axises of ellipse for intrinsic direction information of hand gestures. The gradient of object indicates relative Figure using the combination of moment values in Equation (7).

$$\theta=\frac{1}{2}arctan\frac{2\eta_{11}}{\eta_{20}-\eta_{02}} \tag{7}$$

Table 2 shows the conditions for correcting the angle. In this article, ‘I, J’ in 2-cluster, ‘U, H’ in 4-cluster, ‘G, Z’ in 9-cluster are classified using the direction feature in Table 3.

Table 3. The Range of the Diretion Feature According to Gestures

Gestures	Range of the direction feature
I and J	if$(\theta > 30°)$ $then$ $'J'$ $else$ $'I'$
G and Z	if$(\theta > 34°)$ $then$ $'G'$ $else$ $'Z'$
U and H	if$(\theta > 32°)$ $then$ $'U'$ $else$ $'H'$

4. Experimental Results

4.1. Experimental Environment

The proposed algorithm is implemented in Microsoft foundation class (MFC) based on Visual Studio 2008 C++ and carried out on a PC with a 3.4GHz Intel Core i7 processor with 8GB of memory. Experiment used a Kinect camera with infrared light. This camera has advantages. The first is invariant to lighting changes. The second can easily detect the hand image regardless of the background image due to immediately get depth information through hardware.

4.2. Experimental Results

Figure 9 shows the resulting images of the preprocessing in ASL 'P'. Figure 10 shows the resulting images of ASL recognition. In each Figure, a top right image is RGB image, a bottom right image is hand detection image in infrared image. And five top left images are binary image deleted inner noise, edge detection image, a depth invariants image, an infrared image, and a binary image by turns. Consol window indicates width, height, the number of pixels in hand region, edge length, area, circularity, eccentricity, moment values and angle value. Finally, a bottom center image is recognition resulting image.

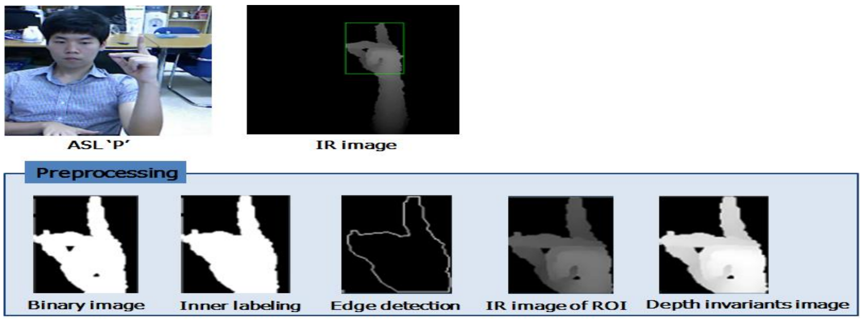
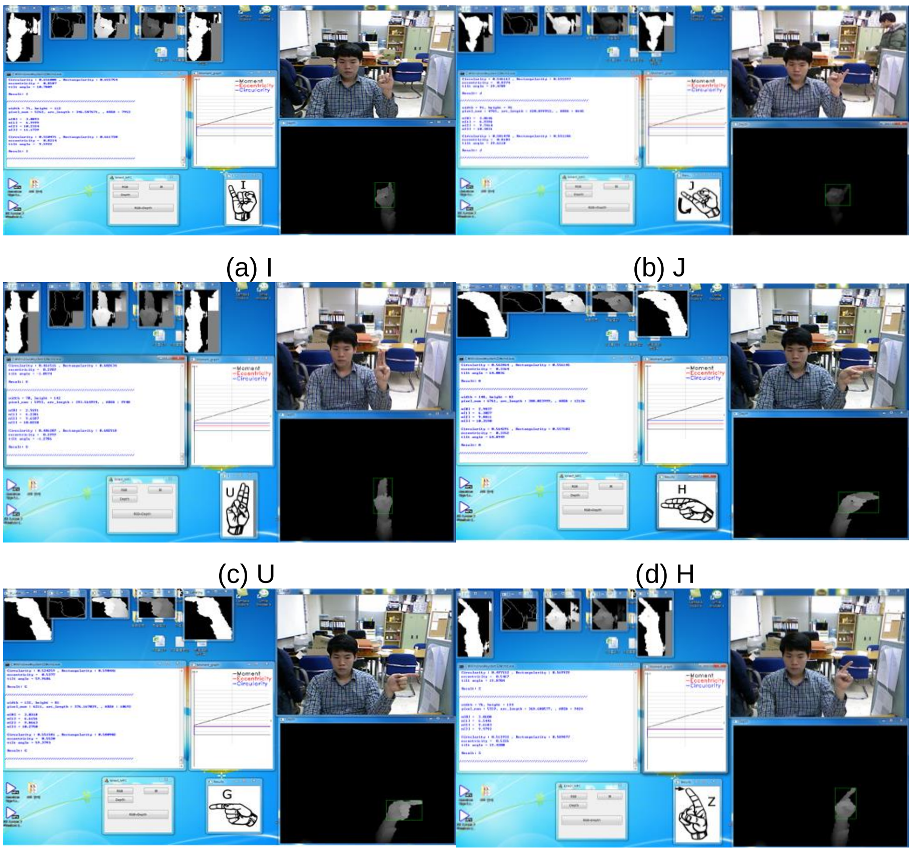


Figure 9. The Resulting Images of the Preprocessing



(a) I (b) J

(c) U (d) H

(e) G (f) Z

Figure 10. The ASL Recognition Results (I, J, U, H, G, Z)

Table 4. The Comparison of the Recognition Rate and Average Processing Time for Each Method

	Methods		
	The method based on contour	The method without direction	The proposed method
Recognition rate(%)	93.81	95.21	97.15
Sec/frame	0.046	0.039	0.041

Table 4 indicates the comparison of the recognition rate and average processing time for each method. The recognition rate of the proposed method is higher than other methods. Through experiments, the proposed method can increase the recognition rate at about 3.34%. In case of the proposed method, it takes 0.041 s to process one frame. Owing to extracting direction information, it is slower than the method without direction. However, the proposed method is faster than the method based on contour and it is sufficient time to recognize the hand gestures in real time.

5. Conclusions

A hand gestures recognition system based on hierarchical features is proposed for accurate recognition of ASL. This article has demonstrated that similar hand gestures can be classified using hierarchical features. In general, the recognition rate of the proposed method is higher than other methods. The average recognition rate for the proposed method is 97.15%. Through experiments, the proposed method can increase the recognition rate about 3.34%. In this article, we improve the recognition rate through the experiments of various gestures. By using the hand gesture recognition system, the application of the complex shape recognition is expected in future.

Acknowledgments

This research was supported by Nurimaru R&BD Project (Busan IT Industry Promotion Agency) in 2015.

References

[1] Hamada Y., Shimada N. and Shirai Y., "Hand Shape Estimation Using Image Transition Network", In: proceedings of workshop on Human Motion, **(2000)**, pp. 161-166.
[2] Ong S.C.W. and Ranganath S., "Automatic sign language analysis: a survey and the future beyond lexical meaning", J. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, **(2005)**, pp. 873-891.
[3] Park H., Shi H.S., Kim H.H. and Park K.H., "A User Adaptation Method for Hand Shape Recognition Using Wrist-Mounted Camera", J. KIECS, vol. 8, no. 6, **(2013)**, pp. 805-814.
[4] Sigal L., Sclaroff S. and Athitsos V., "Skin color-based video segmentation under time-varying illumination", J. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 7, **(2004)**, pp. 862-877.
[5] X. N. Zhang, Z. Jiang, J. L. Liang, Z. H. and Liu C. L., "Skin color enhancement based on favorite skin color in HSV color space", J. IEEE Transactions on Consumer Electronics, vol. 56, no. 3, **(2010)**, pp. 1789-1793.
[6] Phung S. L., Bouzerdoum A. and Chai D., "Skin segmentation using color pixel classification: analysis and comparison", J. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, **(2005)**, pp. 148-154.
[7] Lawgali A., Bouridance M., Angelova M. and Ghassemlooy Z., "Handwritten Arabic Character recognition: Which Feature Extraction Method", IJAST, vol. 34, **(2011)**, pp. 1-8.
[8] Hu M. K., "Visual Pattern Recognition by Moment Invariants", J. IRE Transaction on information theory, vol. 8, no. 2, **(1962)**, pp. 179-187.
[9] Belkasim S. O., Shridhar M. and Ahmadi M., "Pattern recognition with moment invariants: A comparative study and new results", J. Pattern Recognition, vol. 24, no. 12, **(1991)**, pp. 1117-1138.

[10] Kaensar C., "Analysis on the Parameter of Back Propagation Algorithm with Three Weight Adjustment Structure for Hand Written Digit Recognition", In: 10^{th} IEEE Conference on Service Systems and Service Management, IEEE Press, New York, **(2013)**, pp. 18-22.
[11] Ishwarya M. V., "An Improved Online Tamil Character Recognition Using Neural Networks", IJAST, vol. 42, **(2012)**, pp. 1-10.
[12] Pineda F. J., "Generalization of back-propagation to recurrent neural networks", Physical Review Letters, vol. 59, no. 19, **(1987)**, pp. 2229—2232.
[13] Velu C. M., Vivekanadan P. and Kashwan K. R., "Indian Coin Recognition and Sum Counting System of Image Data Mining using Artificial Neural Networks", IJAST, vol. 31, **(2011)**, pp. 67-80.
[14] Wang J. and Tan Y., "Efficient Euclidean distance transform algorithm of binary images in arbitrary dimensions" J. Pattern Recognition, vol. 46, no. 1, **(2013)**, pp. 230-242.
[15] Chen C. C., "Improved moment invariants for shape discrimination", J. Pattern Recognition, vol. 26, no. 5, **(1993)**, pp. 683-686.
[16] Tou J. T. and Gonzalez R. C., "Pattern Recognition Principles", Addison-Wesley; 2nd edition, **(1977)**.
[17] Wagstaff K. and Rogers S., "Constrained K-means clustering with background knowledge", In: Proceedings of the Eighteenth International Conference on Machine Learning (ICML), **(2001)**, pp. 577-584.

