Advanced Science and Technology Letters
Vol.122 (IRTT 2016), pp.89-94
http://dx.doi.org/10.14257/astl.2016.122.17

A Robust Hand Gesture Recognition Using Combined Moment Invariants in Hand Shape

Seungmin Leem[1], Hyeonseok Jeong[1], Yonghwan Lee[2], Sungyoung Kim[1,1],

[1]Dept. Of Computer Engineering, Kumoh National Institute of Techonogy, Daehak-ro 61, Gumi, Gyeongbuk 730-701, Korea
[2]School of Electronics Engineering, Kumoh National Institute of Techonogy, Daehak-ro 61, Gumi, Gyeongbuk 730-701, Korea
{mapsosa215, asdfgh25}@naver.com, {yhlee, sykim}@kumoh.ac.kr

Abstract. In this paper, we suggest a method that recognizes hand gesture based on moment features in hand shape. First of all, hand regions are segmented from input streams based on skin color detection. Hand detection can be achieved more easily if smart devices such as Kinect are used but we used web camera as an input device. Because hand and also face can be segmented from a frame, we try to remove face from the segmented result. From segmented hand regions palm region is extracted by removing wrist and then moment invariants are calculated from the palm region. Finally we use artificial neural network to classify the classes of the hand gestures. We perform recognition test for input patters with trained DB of 7 classes that contains hand gesture of rock-paper-scissors game and 3 different kinds of hand shape concerned with robot control.

Keywords: Hand Gesture, Gesture Recognition, Moment Invariants, Combined Moment Invariants

1 Introduction

Hand detection and hand gesture recognition play key role in multimedia applications such as human-device interaction (HDI). There are some excellent software libraries or camera systems to detect hands. Kinect and Leap motion sensor are typical devices. But smaller and compact camera systems are embedded in another devices such as smart electric whiteboard system.

In smart electric whiteboard system, hand detection is useful for user to write on the board on the position away from a location of the board. Also hand gesture recognition are used to control the whiteboard system at remote location. Kim and Cheon [1] try to separate strokes from hand motion trajectories for smart electric whiteboards. To capture hand motion trajectories, hand detection should be achieved first.

[1] Corresponding author: Dept. of Computer Engineering, Kumoh National Institute of Techonogy, Daehak-ro 61, Gumi, Gyeongbuk 730-701, Korea (sykim@kumoh.ac.kr)

ISSN: 2287-1233 ASTL

There are many kinds of shape descriptors but moment invariants are very classical and powerful ones. A moment is a quantitative measure that capture the shape or significant features of a set of points. We can represent shape information of objects efficiently and compactly with moment invariants. We try to discriminate the gesture of the hands based on moment invariants but the invariants can be used for many other applications that recognize objects such as digits based on their shapes.

In this paper, we suggest a method that detect hands from video sequence and recognizes gestures of the hands based on moment features of the hand shape. First of all, hand regions are segmented from input video streams based on skin color detection. Because hand and also face can be segmented together from an input frame, we try to remove face from the segmented result by Haar cascades [2]. From segmented hand regions palm region is extracted by removing wrist and then moment features are calculated from the palm region. Finally we use artificial neural network to classify the classes of the hand gestures.

2 Moment invariants

As aforementioned, a moment is a quantitative measure that capture the shape or significant features of a set of points. Flusser et al. [3] define "moments are projections of a function onto a polynomial basis similar with Fourier transform". Especially moments on images are calculated based on intensities of image pixels and the projection function. After all moments model the shape of an object mathematically by describing distribution of gray level of images that are normalized.

Because moments describe some characteristics of objects, it is necessary to segment images into significant objects. Moments show mean, variance, skewness, kurtosis (platykurtic or leptokurtic) of the probability density from images. Geometric moments m_{pq} of an image are defined as in Eq. (1) where $r=p+q$ is called an order of the moment.

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

Hu [3] introduced new moment invariants that are invariant to in-place rotation around the origin. Hu moments are also invariant to translation and scaling. Regular moments such as Hu moments have some problems [3]. These problem can be overcome by using orthogonal moments that use orthogonal basis sets. There are some kinds of orthogonal moments such as Zernike, Tchebichef and Krawchouk.

3 Recognition of hand gestures

3.1 Extraction of hand regions

Hand regions are segmented based on skin colors from input sequences in webcam. Skin colors can be segmented stably and accurately if we use the segmentation method such as in [4] and [5]. These methods, however, need computation time to segment and are not adequate for real-time applications.

In this paper, we try to segment skin colors based on predefined range of color components in HSV color space. A segmentation result is seen in Fig. 1. Because faces are sometimes extracted with hands, we must remove the faces from the segmented results. We use a method to detect faces using Haar-like features. Fig. 2 shows the result of removing face and followed by morphological operations.

Fig. 1. This shows a result of skin color detection. We can segment skin color regions properly in indoor environment, as shown in the right image.

Fig. 2. Left image shows a result of removing a face from skin regions. We use a face detection algorithm based on Haar-like feature to remove faces. Right image shows a result after applying morphological operations.

After removing faces from skin color regions, we find a largest component that represents a hand region by pixel-labelling and size-filtering. We want to segment just hand regions but wrists are sometimes segmented together. So we need to localize hand region by separating from wrist regions. To separate a hand region from a wrist, we use distance transform [6] that calculates the distance between foreground pixels and their closest background. We can find most inner position of an object and remove foreground pixels that locate outside of circle centered in the inner position and with predefined radius.

3.2 Classification of hand shapes

After segmenting hand regions, we calculate feature vectors that represent hand shapes from extracted hand regions. Two kinds of moment invariants are used to represent the shapes – Hu and Zernike moment [3]. Then we create a new feature by fusing two existing features. Fusion of features can be achieved by two methods – serial and parallel. We use a serial fusion approach that combines features in sequential order. We use artificial neural network as a classifier.

We use seven kinds of classes of hand shapes to train a classifier. Fig. 3 shows some examples of the classes. Four of classes are for rock-paper-scissors game and the others for robot control. We can control a robot that is built with LEGO Mindstorms.

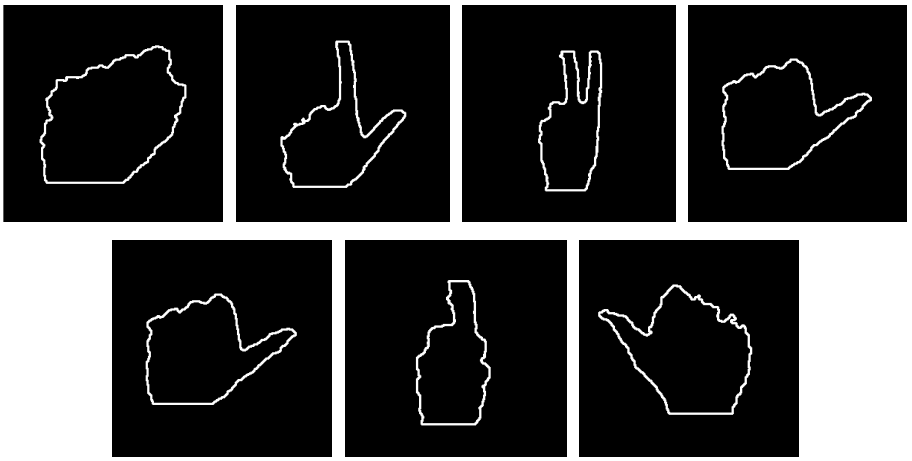

Fig. 3. There are some examples of images that belong classes. We use seven classes to train hand shapes. Four classes among them are for rock-paper-scissors game and the others for robot control.

4 Experimental results

We try to classify an input hand shape into seven hand classes. As aforementioned, four of them are for rock-paper-scissors game. Fig. 4 shows the interface of the game system. We use a webcam of MS Lifecam VX-3000 for experiments. The results show that the moment invariants are robust in rotating and scaling objects. And proposed methods provides accurate results in indoor environment and in real time.

Table 1 shows classification accuracy. For experiments, we use two kinds of moment invariants – Hu and Zernike moments. We also use feature vectors that combine two existing moment invariants. Table 2 shows processing time for each step. As you can see, step of detecting faces is consuming most time against other steps. Also calculation of Zernike moment invariants need some computing time and so we use only contours instead of regions of objects to calculate the invariants.

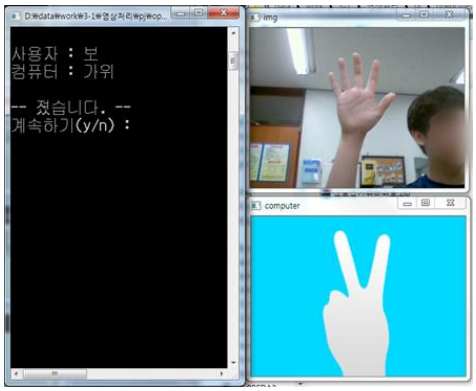

Fig. 4. This shows a system for rock-paper-scissors game. User and a computer system select a pattern of four possible pattern, respectively. Then this system prints result of the winning and losing the game.

Table 1. This show evaluation of the classification result

Class	Hu	Zernike	Hu+Zernike
Rock	97.33	98.00	98.00
Scissors (thumb & index finger)	94.66	96.00	97.33
Scissors(thumb & middle finger)	96.00	98.00	99.33
Paper	99.33	98.00	98.00
Thumb	95.33	96.00	98.00
Thumb to left	86.00	86.66	97.33
Thumb to right	60.66	73.33	91.33
Average	89.56	92.28	97.04

Table 2. This show pprocessing time of each step of classification

step	time
Detecting skin colors	0.022ms
Detecting faces	0.398ms
Labeling	0.002ms
Finding an inner position of a hand	0.002ms
Calculating Hu moments invariants	0.001ms
Calculating Zernike moments invariants	0.112ms

5 Conclusions

We suggested a robust system for hand gesture recognition based on moment invariants. We used two kinds of moment invariants – Hu and Zernike moments. Each moment invariants could provide recognition accuracies to some extend but we improved the accuracy by combining these two moment invariants.

Suggested method can be used for recognition of any kinds of objects that need to classify their classes based on shape. This method would be suitable for digit recognition if we complement features and functions that combine moment invariants.

Acknowledgement. The research was supported by 'Software Convergence Cluster Program', through the Ministry of Science, ICT and Future Planning (2015-GSWC-17)

References

1. Kim, S., and Cheon J.J, Strokes Separation from Hand Motion Trajectories., 4th International Conference on Circuits, Control, Communication, Electricity, Electronics, Energy, System, Signal and Simulation (2014), Vol. 51, no. 1, pp. 64-67
2. Viola, P. and Jones, M, Rapid Object Detection using a Boosted Cascade of Simple Features. Conf. on CVPR (2001).
3. Flusser, T., Suk, T. and Zitova, B, Moments and moment invariants in pattern recognition. Wiley (2009)
4. Jones M.J. and Rehg J.M, Statistical Color Models with Application to Skin Detection. Int. J. of Computer Vision (2002), Vol. 46, No. 1, pp. 81-96
5. Kim, S.Y. and Ko, J.P, Skin Color Detection Using PCA-based Color Representation. International Journal of Multimedia and Ubiquitous Engineering (2014), Vol. 9, No. 10, pp. 231-242
6. Kimmel, R., Kiryati, N. and Bruckstein, A.M, Distance maps and weighted distance transforms. Journal of Mathematical Imaging and Vision, Special Issue on Topology and Geometry in Computer Vision (1996), Vol. 6, pp. 223-233

