Online available since 2013/Jan/25 at www.scientific.net

doi:10.4028/www.scientific.net/AMM.284-287.3004

A Real Time Hand Gesture Recogni-tion System Based on DFT and SVM

Wen-Her Chen [1], Ching-Tang Hsieh [2, a] and Tsun-Te Liu [1]

[1] Dep. of Physical Education,Tamkang University, New Taipei City, Taiwan

[2] Dep. of Electrical Engineering, Tamkang University, Taipei Taiwan

[a]Correspondending Author: hsieh@ee.tku.edu.tw

Keywords: Camshift, BEA, SVM, Hand Gesture Recognition

Abstract

Vision based band gesture recognition provides a more nature and powerful means for human-computer interaction. A fast detection process of hand gesture and an effective feature extraction process are presented. The proposed a hand gesture recognition algorithm comprises four main steps. First use Cam-shift algorithm to track skin color after closing process. Second, in order to extract feature, we use BEA to extract the boundary of the hand. Third, the benefits of Fourier descriptor are invariance to the starting point of the boundary, deformation, and rotation, and therefore transform the starting point of the boundary by Fourier transformation. Finally, outline feature for the nonlinear non-separable type of data was classified by using SVM. Experimental results showed the accuracy is 93.4% in average and demonstrated the feasibility of proposed system.

1. Introduction

The computer is becoming more pervasive in contemporary life. Hand gesture recognition is a hopeful research field and has gained much attention in computer vision. Its most application is effective and friendly interfaces between human and machine known as human-computer interaction (HCI). In addition, it can be used for teleconferencing, interpretation and learning of sign languages because it does not need expensive and special hardware. Hand gesture recognition methods are have already been presented, which can be mainly divided into “Data glove-based” and “Vision-based” methods. These Data glove-based method [1-4] is use sensor devices for digitizing hand, but it is annoying to use. Thus there recently has been an increasing number of gesture recognition technique using vision-based methods. Although, it is a very challenging interdisciplinary project in order to the following three reasons [5]: hand gestures are rich in diversities, multi-meanings, and space–time varieties; human hand is a complex non-rigid object; computer vision itself is an ill-posed problem. In general, the approach of vision-based hand gesture recognition can be divided into two categories, one based on 3D hand / arm model [6] [7], the other based on appearance modeling. The methods based on 3D hand / arm model are favorable for communicative hand gestures and often use in animation. But, it is not only the highly computational complexity but also fallibility. Then the methods based on appearance model are matching really on the colors of the objects and there are many advantages. For example, low computational complexity, real-time process, and so forth.
Burger et al [8] propose a demonstrator on Cued Speech hand gesture recognition. Chen et al [9] present an approach to recognize for the robot vision system and this method comprises three main steps. The first analyzes the distribution of the specific color in hue saturation intensity HSV color space in order to detect hand gestures. The second uses fourier trans-formation to generate a set of complex numbers. Finally, out-line feature was classified by using SVM. These methods are robot vision system, but must wear a one coloured glove. Furthermore, it isn’t nature and powerful means for human-computer interaction. Buluswar and Draper [10] present an approach based on automatic target recognition. It uses multivariate decision trees for recursive non-parametric function approximation to learn the color of a target from training samples. Lee et al [11] propose a method to recognize hand gestures extracted from images with complex background. Authors utilize image processing to rely upon recognition through hand extraction by colors. Recently, there are more studies using support vector machines (SVM) to classify, and can get highly accurate systems capable of recognizing gestures [9] [12-14].

In this paper, we use HSV color space to detect hand gestures and Camshift algorithm to track. In recognize hand gestures, first must extract the boundary of the object by Boundary Extraction Algorithm (BEA). Second utilize Fourier transformation to generate shape description. Finally classify shape description by SVM.

This paper is organized as follows. In the next section, we refer to our proposed method. Section 3 reports the experimental results. Finally, conclusion is shown in section 4.

2. Proposed Method

The proposed a hand gesture recognition algorithm comprises five main steps. First, subtract capture image from background after building the background. Next, use Camshift algorithm to track skin color after closing process. In order to extract feature, we use BEA to extract the boundary of the hand. The benefits of Fourier descriptor are invariance to the starting point of the boundary, deformation, and rotation, and therefore transform the starting point of the boundary by Fourier transformation. Finally, outline feature for the nonlinear non-separable type of data was classified by using SVM.

2.1 Camshift Algorithm

This paper [15] applied Mean-shift analysis to the face tracking, in which the color histogram is used as cue. Mean shift algorithm is a general nonparametric clustering method based on density estimation for the analysis of complex feature spaces. The method comprises four main steps. First, use the centroid of the search window to choose a search windows size in the color probability distribution image. Next, derive the centroid of the search window. Find the zeroth moment in Eq. (1) and the first moment for x and y in Eq.(2)(3). Center the search window at mean location in Eq. (4). Where I(x,y) is the pixel value at position (x,y), and x and y range over the search window. Third, reset the size of the search windows in Eq.(5). Finally, repeat steps 2 and 3 until convergence occurs. However, the Camshift algorithm [16] has been used to adaptively meet location of the search windows and the size. Current frame's localization results of the tracked object are reported and used to set the size and location of the search window in the next frame image. The Camshift algorithm comprises five main steps. First, choose the initial location and size of the search window. Next, compute the color probability distribution within the search window. Third, mean shift as above, and get the mean location and new size of the search window. Center the search window in successive frame image at the mean location obtained in Step 3, and repeat Step 3. Finally, calculate orientation and scale of the target. The second moment is showed in Eq. (6-8) and the length l and width w is showed in Eq. (9)(10). Where $a = {M_{20}}/{M_{00}} - x_c^2$, $b = {M_{11}}/{M_{00}} - x_c y_c$, $c = {M_{02}}/{M_{00}} - y_c^2$. Then the direct angle Θ from target length to horizontal is showed in Eq. (11).

Camshift algorithm is quite effective in hand tracking, thus we use this method to track. First, build background and capture one frame of video. Second, detect moving objects by background subtraction. We use close operation to denoising for detecting moving objects. Finally, hand track by Camshift algorithm.

$$M_{00} = \sum_x \sum_y I(x,y) \quad (1)$$

$$M_{10} = \sum_x \sum_y xI(x,y) \quad (2)$$

$$M_{10} = \sum_x \sum_y xI(x,y) \quad (3)$$

$$x_c = {M_{10}}/{M_{00}},\ y_c = {M_{01}}/{M_{00}} \quad (4)$$

$$s = 2\sqrt{M_{00}/256} \quad (5) \qquad s = 2\sqrt{M_{00}/256} \quad (6)$$

$$M_{02} = \sum_x \sum_y y^2 I(x,y) \quad (7) \qquad M_{11} = \sum_x \sum_y xyI(x,y) \quad (8)$$

$$l = \sqrt{\left[(a+c) + \sqrt{b^2 + (a-c)^2}\right]/2} \quad (9) \qquad w = \sqrt{\left[(a+c) - \sqrt{b^2 + (a-c)^2}\right]/2} \quad (10)$$

$$\Theta = \frac{1}{2}\arctan\left[\frac{2\left(M_{11}/M_{00} - x_c y_c\right)}{\left(M_{20}/M_{00} - x_c^2\right) - \left(M_{02}/M_{00} - y_c^2\right)}\right] \quad (11)$$

2.2 Boundary Extraction Algorithm

We extract the boundary of the hand in order to recognize hand gestures in the tracking image. This method is proposed by Liu [17] and comprises four main steps. First, scan the binary image from top to bottom and left to right to find the first 255 pixel as the starting point and record the position into x and y arrays, i.e., x[1]=I and y[1]=J. And let the direction number of the starting pixel be 0. Next, we assume that I is the direction number of current pixel and J is the direction number for searching next pixel and define a decision rule for searching next pixel as: (I,J)={(0,6),(1,7),(2,0),(3,1),(4,2),(5,3),(6,4),(7,5)}. Third, scan clockwise until the first pixel value which is 255, and record its position into x and y arrays. Finally, repeat steps 2 to 4 until the first pixel value which is the same as the starting pixel.

2.3 Fourier Descriptors

Cosgriff [18] proposed Fourier transformation of a boundary function and generates a set of complex numbers which are called Fourier descriptors. Because the benefits of Fourier descriptor are invariance to the starting point of the boundary, deformation, and rotation, we use Fourier descriptor to shape description based on the boundary of the object.

The x and y from boundary extraction algorithm operation can be represented as curve functions a[L] and b[L], where L is the number of pixels in the boundary. The discrete Fourier transforms of a[L] and y[L] are showed in Eq.(12). Where w=2π/L, a[n] and b[n] are the coefficients of the FDs, i.e., s[n]=a[n]+jb[n] called the FDs of the boundary. Finally, we have to normalize s[n] in Eq.(13)(14).

$$a[n] = \frac{1}{L}\sum_{m=1}^{L} x[m]e^{-jn\omega_0 m} \qquad b[n] = \frac{1}{L}\sum_{m=1}^{L} y[m]e^{-jn\omega_0 m} \quad (12)$$

$$r[n] = \sqrt{|a[n]|^2 + |b[n]|^2} \quad (13)$$

$$s[n] = \frac{r[n]}{r[1]},\ \ s[1] \equiv 1 \quad (14)$$

$$yi(w \times xi + b) > 0, i = 1, \ldots, N \quad (15)$$

$$yi(w \times xi + b) \geq 0 \quad (16)$$

$$yi(w \times xi + b) \geq 1 - \varepsilon, i = 1, \ldots, N \quad (17)$$

$$k(x,y) = (\emptyset(x), \emptyset(y)) \quad (18)$$

2.4 Support Vector Machine

Support Vector Machines (SVM) [19] is a linear machine and a powerful tool for resolving classification problems. This method's aim is to create a hyper-plane as the decision surface. The SVM exploited statistical learning theory to know a minimum upper bound on the generalization error.

A. Optimal Separating Hyper-plane

The Optimal Separating Hyper-plane can find the maximal margin of separation. We assume the class is represented by the subset Di=+1 and Di=-1 are linearly separable and then we find the weight vector w and bias is showed in Eq.(15). If there is a hyper-plane that can satisfy Eq. (15) then we call it linearly separable. Once a linearly separable hyper-plane is obtained and is possible to rescale the bias b and the weight w that the closest point to the hyper-plane has a distance of 1/||w||. Eq. (15) can be showed as Eq. (16). The one with the distance to the closest point is maximal that is called the optimal separating hyper-plane (OSH).

B. Linearly non-separable case

Linearly non-separable case is not linearly separable. And it is impossible to create a hyper-plane that can separate the classes and there are any errors as coming across any classification. Then it is possible to find an optimal hyper-plane that minimizes the probability of the classification. We introduce a slack variable that measures deviation of a data point from ideal position of linear as Eq. (17).

C. Nonlinear Support Vector Machines

The Pattern classification is made up of a multiclass nonlinear set of input data thus Nonlinear Support Vector Machines must be implemented. This happens that converting the nonlinear input space to a high-dimensional linear feature pace by a nonlinear mapping. Therefore, replace x with its feature space mapping so as Eq. (18). This is evaluated using dot product, but can be computational large. Therefore, it is important to choice of kernel k and can be reduced large quadratic computation. In this paper, in order to recognize hand gestures posed at different angles, multiple images were acquired from camera.

$$\mathrm{yi}(\mathrm{w} \times \mathrm{xi} + \mathrm{b}) > 0, i = 1, \dots, \mathrm{N} \quad (15)$$

$$\mathrm{yi}(\mathrm{w} \times \mathrm{xi} + \mathrm{b}) \geq 0 \quad (16)$$

$$\mathrm{yi}(\mathrm{w} \times \mathrm{xi} + \mathrm{b}) \geq 1 - \varepsilon, \mathrm{i} = 1, \dots, \mathrm{N} \quad (17)$$

$$\mathrm{k}(\mathrm{x}, \mathrm{y}) = (\emptyset(\mathrm{x}), \emptyset(\mathrm{y})) \quad (18)$$

Table 1. Recognition rate of the system

Gesture	1	2	3	4	5
$\frac{\text{number of frame}}{\text{total}}$	$\frac{1920}{2000}$	$\frac{1880}{2000}$	$\frac{1870}{2000}$	$\frac{1860}{2000}$	$\frac{1810}{2000}$
Recognition Rate	96%	94%	93.5%	93%	90.5%

3. Experimental Results

The hand gesture region detection and recognition approach was implemented on an Intel Core 2 Duo E7200 2533 MHz running Window XP with an USB Logitech web camera. All tests were executed on 320×240 images with 30b color depth, which were sampled at a rate of 30 frames/s. The recognition method can be used with real world images as the input mechanism of a user interface, surveillance system, autonomous robot vision system or virtual environment applications.

In order to recognize hand gestures posed at different angles, multiple images were acquired from camera. Each gesture contains 1000 images and a total of 5000 images were taken for all the gestures. The results of recognition rate of the system for testing data are shown in Table 1.

4. CONCLUSION

A fast detection process of hand gesture and an effective feature extraction process combined Camshift, BEA, DFT with SVM algorithm are presented. First use Camshift algorithm to track skin color after closing process. Second, in order to extract feature, we use BEA to extract the boundary of the hand. Third, the benefits of Fourier descriptor are invariance to the starting point of the boundary, deformation, and rotation, and therefore transform the starting point of the boundary by Fourier transformation. Finally, outline feature for the nonlinear non-separable type of data was classified by using SVM. Experimental results showed the accuracy is 93.4% in average and demonstrated the feasibility of proposed system.

References

[1] Y. Mikio, Interface system based on hand gestures and verbal expressions for 3-D shape generation, Terebijon Gakkaishi/Journal of the Institute of Television Engineers of Japan, vol.50, no.10, pp.1482-1488, 1996.

[2] V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction : A Review," IEEE Trans. PAMI., vol. 19, no. 7, 1997.

[3] C. Oh-Young, K. Hyoung-Gon, K. Sung-Jea and A. Sang Chul, A Hand Gesture Recognition System for Interactive Virtual Environment, IEEK, 36-s(4), pp. 70-82, 1999.

[4] T. S. Huang and V. I. Pavlovic, Hand Gesture Modeling, Analysis and Synthesis, Proc. Int'l Workshop on Automatic Face and Gesture Recognition, pp. 73-79, 1995.

[5] Teng, X., Wu, B. , Yu, W., Liu, C. : A hand gesture recognition system based on local linear embedding. Journal of Visual Languages and Computing, vol.16, pp.442–454, 2005.

[6] N. Hiroaki, Virtual environment for modeling 3D objects through spatial interaction, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics 6 (1999) VI-81–VI-86.

[7] S. Mu-Chun, A neural-network-based approach to recognizing 3D arm movements, Biomedical Engineering—Applications, Basis and Communications, vol.15, no.1, pp.17-26, 2003.

[8] T. Burger, A. Caplier, and S. Mancini, Cued speech hand gestures recognition tool, in Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05), Antalya, Turkey, September 2005.

[9] C. Kuan-Yu, C. Cheng-Chin, C. Wen-Lung and T. Jyh-Tong, An Integrated Color and Hand Gesture Recognition Approach for an Autonomous Mobile Robot, 3rd International Congress on Image and Signal Processing (CISP), pp. 2496 – 2500, 2010.

[10] S. D. Buluswar and B. A. Draper, Learning object color: ATR in color images, 8th British Machine Vision Conference, Sept. 1997.

[11] JongShill Lee, YoungJoo Lee and EungHyuk Lee, Hand region extraction and Gesture recognition from video stream with complex background through entropy analysis, 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1513-1516, 2004.

[12] C. Yen-Ting and T. Kuo-Tsung, Multiple-angle Hand Gesture Recognition by Fusing SVM Classifiers, IEEE International Conference on Automation Science and Engineering, pp.527-530, 2007.

[13] Yu Ren and Fengming Zhang, Hand Gesture Recognition Based on MEB-SVM, International Conference on Embedded Software and Systems, pp. 344-349, 2009.

[14] H. Deng-Yuan, H. Wu-Chih and C. Sung-Hsiang, Vision-Based Hand Gesture Recognition Using PCA+Gabor Filters and SVM, 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1513-1516, 2009.

[15] Gao, Jianpo;Wu, Zhenyang; Wang, Yujian. Face tracking algorithm based on mean shift and ellipse fitting. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v 4233 LNCS-II, 2006, p 270~277.

[16] Bradski, G. R. "Computer Vision Face Tracking for Use in a Perceptual User Interface". Intel Technology Journal, vol.2, no.2, pp13-27, 1998.

[17] Y. H. Liu, Feature Analysis and Classifier Design and Their Applications to Pattern Recognition and Data Mining, Ph.D. Disseration, National Taiwan Univ., R.O.C., 2003.

[18] R. L. Cosgriff, Identification of shape, Technical Report 802-11, Ohio State University, Columbus, OH, Dec. 1960.

[19] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, pp. 82-108, Cambridge University Press, 2004.

Innovation for Applied Science and Technology
10.4028/www.scientific.net/AMM.284-287

A Real Time Hand Gesture Recognition System Based on DFT and SVM
10.4028/www.scientific.net/AMM.284-287.3004