Indonesian Journal of Electrical Engineering and Computer Science
Vol. 8, No. 2, November 2017, pp. 482 ~ 486
DOI: 10.11591/ijeecs.v8.i2.pp482-486

Fingertip Detection Using Histogram of Gradients and Support Vector Machine

Ali Sophian, Dayang Qurratu'aini
Department of Mechatronics Engineering, Faculty of Engineering, International Islamic University Malaysia
Jalan Gombak, 53100, Kuala Lumpur, Malaysia
*Corresponding author, e-mail: ali_sophian@iium.edu.my

Abstract

One important application in computer vision is detection of objects. This paper discusses detection of fingertips by using Histogram of Gradients (HOG) as the feature descriptor and Support Vector Machines (SVM) as the classifier. The SVM is trained to produce a classifier that is able to distinguish whether an image contains a fingertip or not. A total of 4200 images were collected by using a commercial-grade webcam, consisting of 2100 fingertip images and 2100 non-fingertip images, were used in the experiment. Our work evaluates the performance of the fingertip detection and the effects of the cell's size of the HOG and the number of the training data have been studied. It has been found that as expected, the performance of the detection is improved as the number of training data is increased. Additionally, it has also been observed that the 10 x 10 size gives the best results in terms of accuracy in the detection. The highest classification accuracy obtained was less than 90%, which is thought mainly due to the changing orientation of the fingertip and quality of the images.

Keywords: *fingertip detection, machine vision, support vector machine, histogram of gradients*

1. Introduction

The application of machine vision has served various sectors, such as manufacturing, transportation, education, and health. This has been made possible by the affordability of the vision systems and processing power. New opportunities have, therefore, arisen which are waiting to be exploited in order to improve human wellbeing. One important area to be explored is health applications that can be used in the treatment of patients who need physical therapies to restore their normal physical activities.

There is also high potential of using machine vision to help hand rehabilitation for patients who have survived stroke. Some of these patients are required to squeeze a therapy ball repetitively as part of the therapy [1]. In order to make allow more objective evaluatin on the progress, one of the challenges is to have the ability of detecting the fingers to allow some non-contact measurement of the finger movements. Therefore, visual object detection is thought to be potentially useful for such applications. There are many possible tecniques to be applied. This works presents and evaluates the performance of a detection technique that uses histogram of gradients (HOG) as the features and support vector machine (SVM) as the classification technique. The combination of HOG and SVM can generate classification accuracy higher than 94% in various applications [2], [3].

It should be noted that the problem faced in this work is rather different from many fingertip detection problems reported in various works as most of them deal with extended fingers and no occlusion is present, whilst in this problem the fingers are bent and they are generally partially occluded, including the most part of the hand. It is desirable that the patients do not have to put on any sensors or gloves, whose preparation may demotivate them to perform the exercise. Therefore, the use of machine vision is considered to be one of the potential candidates for this application thanks to their non-contact nature.

The challenge in this project is even more demanding as the solution is expected to be cost effective so that they can be made available to homes that the patients can use them at the comfort of their homes. This has made the use of consumer-grade webcams attractive although the quality of the images will not be as sharp as industry-grade cameras.

Received June 29, 2017; Revised September 2, 2017; Accepted September 19, 2017

The use of more costly MS Kinect for fingertip detection has been discussed in many literature for gesture detection, such as [4–6]. While the use of Kinect with its depth image is so effective for detection and positioning, the associated cost is deemed prohibitive for general home applications. Besides, its operating minimum distance, which is 1.8 m, is too large for this hand-therapy application. Some approaches need markers for the fingertip recognition [2], which should be avoided for our project as they would require more time from the users to use the system. Many proposed approaches use skin color [7], [8], which can be ineffective when the color of the ball especially is too similar to the skin.

The next section will describe the experimental setup, the data collection. Subsequently, the detection method is briefly described. Then, the results are presented and discussed. Finally, the conclusions that can be drawn from the work are presented.

2. Research Method

2.1. Experimental Setup

This experiment uses consumer-grade Logitech C615 webcam that has a resolution of 1920 x 1080 pixels. The webcam comes has an autofocus and is able to capture images up to 8 megapixels. Figure 1 shows an example of image of a hand holding therapy ball. The subject (hand holding the therapy ball) is facing towards the webcam which is located below the subject as shown in Figure 1. The background of the subject has a light emitting source that contributes to the variety of light intensity in each captured image.

Figure 1. Example of image of therapy-ball holding hand

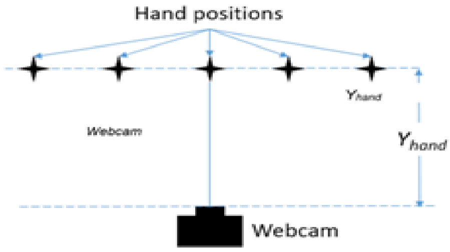


Figure 2. Experimental Setup

Figure 2 shows the webcam setup for the experiment. The stars indicate the positions of the hand where the distance between each two consecutive position is approximately 100 mm from each other. Y_{hand} denotes the perpendicular distance of the position of hands to the webcam, which is also approximately 100 mm. The webcam captures the images of the hand with different orientations at each position.

2.2. Data Collection

Training and testing images were captured from ten individuals, who were comprised of five male and five female adults. Different hand sizes, skin colours, and orientations were represented in the collected image dataset. Figure 3 shows the examples of the hand images of different orientations that have been captured.

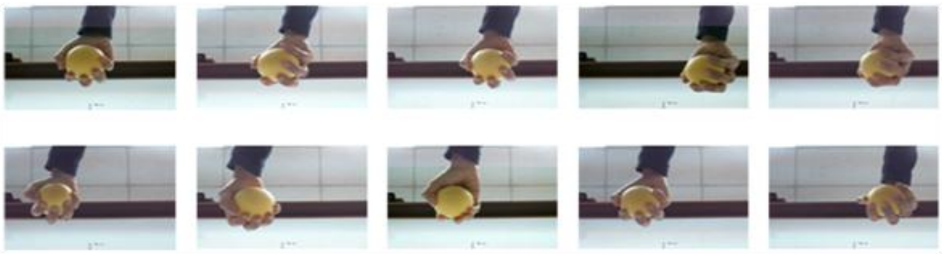


Figure 3. Variation of orientations of the hands due to hand's high degrees of freedom

Following the collection of the images, the images of the fingertip were cropped from the hand holding ball images. The size of each cropped image is 50 x 50 pixels. Furthermore, a set of cropped 50 x 50 images of non-fingertip, which contains the background, wrist, ceiling, etc, were also created from the collected images. A total of 4200 images were created for the work reported in this paper. Examples of the cropped images are shown in Figure 4.

(a) (b)

Figure 4. Examples of the cropped images, (a) Fingertip, (b) Non-fingertip

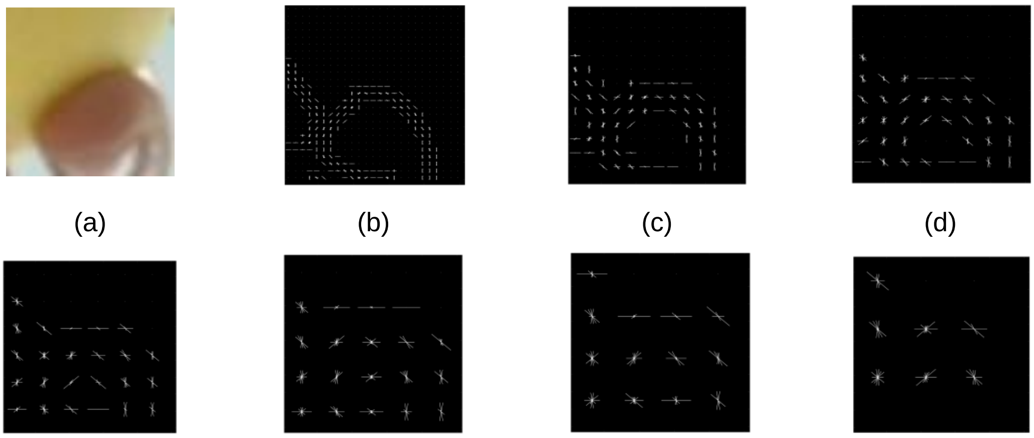
3. Fingertip Detection Algorithm

The algorithm for this report has been implemented in Matlab [2], which poses an extensive library of functions for image processing or machine vision.

3.1 Histogram of Oriented Gradients

Histogram of Oriented Gradients (HOG) is one of wellknown feature descriptors that is used for detection of object in machine vision [9]. With HOG, the distribution or histogram of the gradients of the image is determined, which is then used as features. In images, gradients capture the corners and edges, which are generally useful for detection of object. Gradients can be computed in different ways, including by using some filter kernels.

Normalization is normally performed in order to reduce the effects of varying illumination. Different sizes of cell can be employed for the calculation of the histogram. Figure 5 shows the visualization of the HOG features for different cell sizes for the fingertip image shown in Figure 5(a). The smaller the size of the cell, the higher the dimensionality of the features.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5. (a) An example of an image of a fingertip, (b)-(h) show the HOG features for cell sizes of 2 x 2, 4 x 4 , 6 x 6, 8 x 8, 10 x 10, 12 x 12 and 14 x 14 respectively

3.2 Classification

Following the feature extraction using HOG, classification is perfomerd in order to determine whether the image is containing a fingertip or not. Support Vector Machine (SVM) is used in this work for the classification.

Support Vector Machine (SVM) is a supervised learning method that is used for classification [10]. It carries out classification by creating a multi-dimensional hyperplane which divides the data into groups optimally. This makes SVM classifier model closely associated with neural networks. The SVM classifier model uses a sigmoid kernel function, which is similar to the two-layer perceptron of neural network.

By using randomly chosen dataset, the classifier is trained. Figure 6 shows the steps taken for training the SVM classifier with HOG features as the input. Prior to extraction of the HOG features from the images, they are converted to grayscale and then binary images, i.e. black and white images.

Figure 6. Training Procedure

4. Experimental Results and Discussion

In this section, the accuracy of the detection algorithm will be assessed by using image data sets collected and the generated confusion matrices. Firslty, the effects of the size of the training data will be studied. Then, the effects of the cell size will be investigated in order to find the optimized solution for this study.

How the size of the training data affects the classification accuracy has been plotted in Figure 7. The cell size used in this test is 8x8 The plots show that the classification accuracy is generally improving with larger training data set size for both fingertip and non-fingertip recognition. Considering the trends, it is possible that the results can be improved further if the data sets are increased in size.

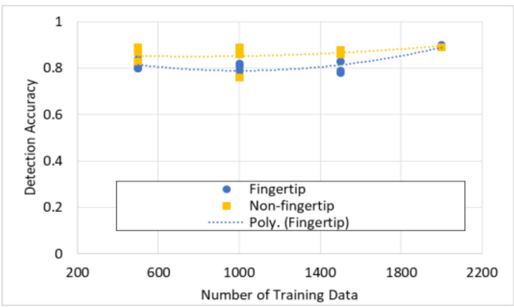


Figure 7. Detection accuracy vs number of training data

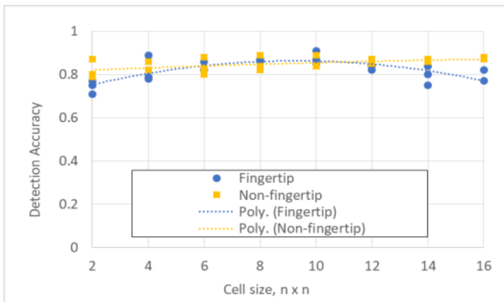


Figure 8. Detection accuracy vs the size of cell, the numbers in the x-axis is the size of the square cell, e.g. 2 means 2 x 2.

Figure 8 shows the plots of the classification accuracy vs the size of the cell. The sizes investigated are 2 x 2, 4 x 4, up to 16 x 16. While the classification accuracy for the non-fingertip is increasing steadily with the increasing size of the cell, interestingly the best performance for detecting the fingertip is found when the size of the cell is between 8 x 8 and 10 x 10. Overall, this would suggest that the size 10 x 10 should be chosen as this would give the highest classification accuracy for the correct identification of both fingertips and non-fingertips. In terms of dimensionality, it would also be good as can be seen from Figure 5-6.

4. Experimental Results and Discussion

In this work, the use of HOG and SVM in the detection of fingertip in the images obtained by using a consumer-grade webcam has been presented. Experimental tests have been obtained, plotted and discussed. As expected, the performance of the detection is improved as the number of training data is increased. The more data we use, the better the performance, although it's likely to hit saturation at one point. As for the size of the cell for the HOG, it was found that the 10 x 10 size would give the best results in terms of accuracy in the detection. However, even at its best, the classification accuracy obtained was less than 90%. This suggests more robust features for this problem maybe needed.

References

[1] R Jaber, DJ Hewson, J Duchene. Design and validation of the Grip-ball for measurement of hand grip strength. *Med. Eng. Phys.* 2012; 34(9): 1356–1361.

[2] K Dorfmuller-Ulhaas, D. Schmalstieg. *Finger tracking for interaction in augmented environments*. in Proceedings IEEE and ACM International Symposium on Augmented Reality, 2001, pp. 55–64.

[3] D-Y Huang, C-H Chen, T-Y Chen, W-C. Hu, K-W Feng. Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads. *J. Vis. Commun. Image Represent.* 2017; 46 Supplement C: 250–259.

[4] JL Raheja, A Chaudhary, K Singal. *Tracking of Fingertips and Centers of Palm Using KINECT*. in 2011 Third International Conference on Computational Intelligence, Modelling & Simulation. 2011: 248–252.

[5] Z Lai, Z Yao, C Wang, H Liang, H Chen, W Xia. Fingertips detection and hand gesture recognition based on discrete curve evolution with a kinect sensor. in *VCIP 2016 - 30th Anniversary of Visual Communication and Image Processing*. 2017.

[6] J Shin, CM Kim. Non-Touch Character Input System Based on Hand Tapping Gestures Using Kinect Sensor. *IEEE Access*. 2017; 5: 10496–10505.

[7] Y Liao, Y Zhou, H Zhou, Z Liang. *Fingertips Detection Algorithm Based on Skin Colour Filtering and Distance Transformation*. 2012 12th International Conference on Quality Software. 2012: 276–281.

[8] A Sophian, DQ Awang Za'aba, W Sediono, H Md Yusof, S Sudirman. Markerless detection of fingertips of object-manipulating hand. *ARPN J. Eng. Appl. Sci.* 2016; 11(23): 13716–13720.

[9] N Dalal, B Triggs. Histograms of oriented gradients for human detection. in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005; 1: 886–893.

[10] NH Dardas, ND Georganas. Real-time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques.pdf. *IEEE Trans. Instrum. Meas.*, 2011; 60(11): 3592–3607.