Hand Gesture Detection and Classification Using Boosted Classifier

Abdul Manan Ahmad[1], Sami M Halawani[2]
[1]Fakulti Sains Komputer dan Sistem Maklumat
Universiti Teknologi Malaysia
Malaysia
[2]Faculty of Computing & Information Technology at Rabigh
King Abdul Aziz University
Saudi Arabia
manan@utm.my, dr.halawani@gmail.com

ABSTRACT: *There are many approaches to detect hand gestures. To narrow the search region, most existing methods construct fixed postures. However, under natural condition it is not realistic to make a user stick on a certain pose. In this paper a method to quickly detect hand shape quickly and robustly is experimented. Firstly, a skin color detector will be utilized to detect the presence of a hand in the image, then a set of image is clustered using k-mediod algorithm. Next, a tree structure of boosted cascades will be constructed. A general hand detector is provided by the main tree while the branches will classify valid shapes. Our experiments show a detection rate of 80% and from that 85% is achieved for recognition.*

Keywords: Gesture Detection, Postures Classification, Color Detector, Image Clustering, M-mediod Algorithm, Feature Classificaiton

Received: 1 July 2013, Revised 27 July 2013, Accepted 2 August 2013

1. Introduction

This paper proposed a framework for detecting and recognizing hand shapes that is user independent and background independent. Due to the complexity of hands which contains 14 joints, this area of research has been an interesting area for a long time. Many previous researchers had been limiting the complexity of their system such as fixing the background, using marked gloves and else. However, to avoid those limited functionality there are also researchers using plenty of method to detect skin colors. In our project, we use a boosted classifiers cascade [9] to detect shape in a grey scale image. A tree structure of boosted cascades classifiers is used as detector where the main tree is used to find all possible hand hypotheses on the image. The matching hypothesis is then go to the branches where more detailed cascades are defined to detect specific hand gestures before determining the exact and shape in that image. The data set however needed to be clustered into identical shapes using a k-mediod clustering on the training image. Section 2 will be detailing on the boosting method for the hand detector tree. Section 3 shows the operation of unsupervised learning on the hand shapes database. Section 4 is experimental discussion and section 5 is the conclusion about this paper.

2. Skin Color Detection

In order to make the system faster, we employ a stage where firstly the skin color in the image presents. This will help determine whether to go to next level of boosting learning or skip the current image and scan the next image. The technique used to detect

the presence of skin tone color by [15]. The reason why we chose to employ this method is because luminance issue is taken care of and thus the detection of the signer's hand is better under any light conditions.

3. Boosting Learning

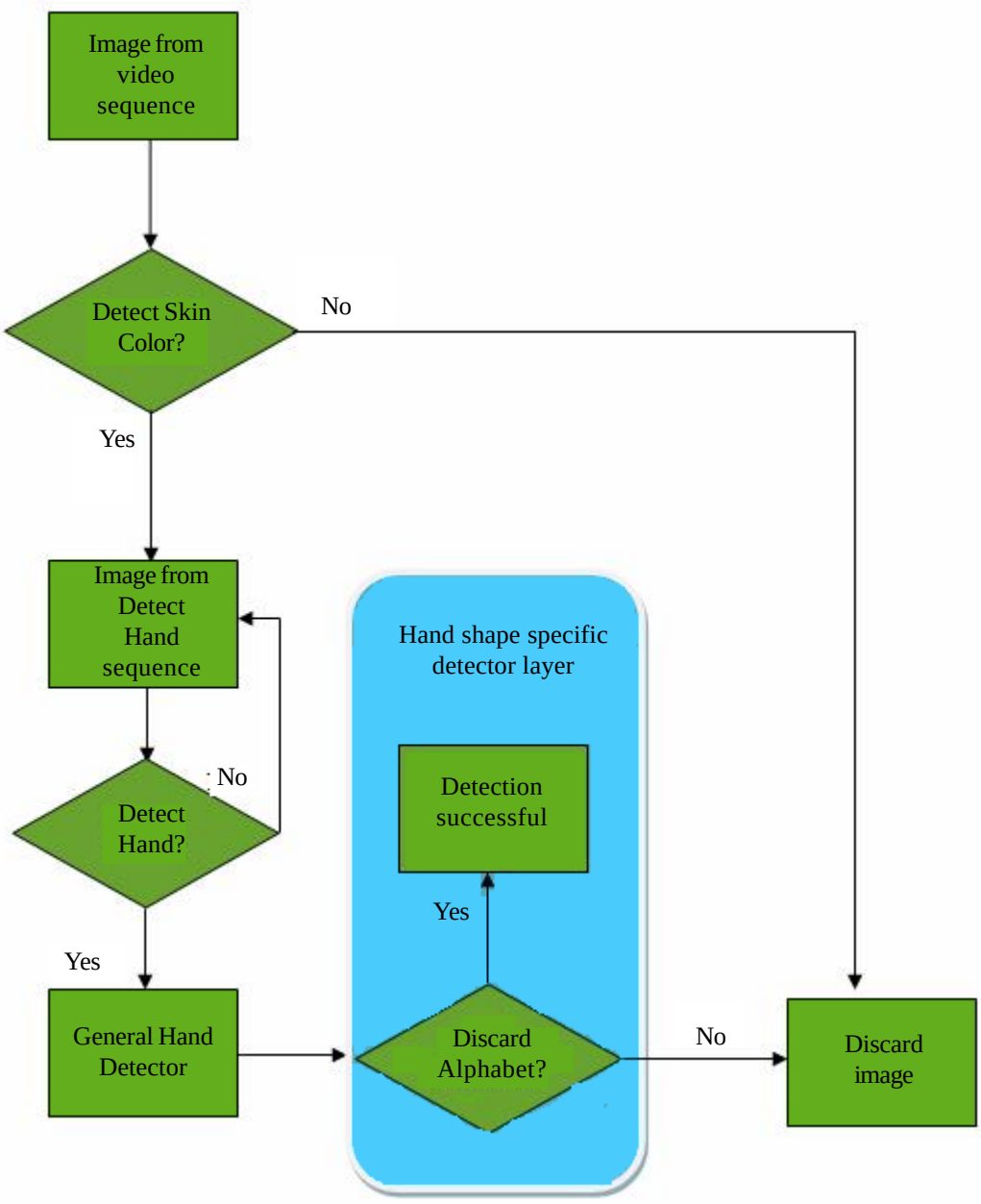

Figure 1. Hand Detector Tree

In [5] boosting is a method to improve the accuracy for a certain learning algorithm. It works by combining weak classifiers to form a strong classifier. Generally, the concept is that a weak classifier performances is somewhat better that the other weak classifiers. However, the weak classifier is used in this paper is to simply detect basic differences in image block efficiently as had been proposed by [9]. Figure 1 is the framework for the hand detectors tree.

Let an image as (i) and (S_C (i)) as strong classifier that defines as signed linear combinations of k (numbers of) weak classifiers (W_c):

$$S_C(i) = sign\left(\sum_{k=0}^{k} W_c\right)$$

As soon as the system detected hands, the value of S_C should be more than zero and less than zero for other object.

3.1 AdaBoost

This algorithm is a machine learning algorithm that functions to improve other algorithms' performance. Due to its adjustability in tweaking the built classifiers, it is also prone to noisy and outliers. It functions by replicating calling the weak classifiers in sets. To use this algorithm in the training phase, we adapt the technique by Ong [14] that is to model the upper bound using an exponential loss function.

$$J(S_c) = \sum_{k}^{N_H} e^{\alpha}$$

(h_i) is the training set of N_H images (not separated yet between hands and unknown objects). The hands image are labeled with $[(y_i) + 1]$ and $[(y_i) - 1]$ for unknown objects image. As we can see from the algorithm, the weak classifiers will be summed sequentially into the set of strong classifiers (S_c) of other classifiers if this upper bound is seen to be decreasing.

Figure 2. Grouped hand gesture from the training images. Number 1 (left) and letter A (right)

3.2 Float Boost

Li et al [13] introduced Float Boost algorithm to construct a strong face-nonface classifier. This algorithm takes the idea of floating search in Ada boost, and then produces similar but higher accuracy in classification than AdaBoost but with less number of weak classifiers. In easy words, it removes weak classifier that does not have any effect in recognition.

3.3 Detection Tree

As in Figure 1, by dividing the last detector to a strong classifier layers, the classification is more practical [9]. So, in the process of detecting hand a detection tree structures is possible to be utilized as suggested in [13] to detect face. This tree structure is made of general detector at the first layer consisting of branch node that contains specific detectors at further deeper layers as in Figure 1 in appendix A. In the other hand, the deeper layer of the tree nodes are trained with more detailed image sets. As the scope is the hand image, the grouping of an image is non-trivial comparing to detecting face.

Due to thousands of images needed to be handled in the system, an automatic method for grouping will be adopted to save time. In this project, we are going to opt for an unsupervised clustering in grouping similar hand shape using k-mediod algorithm as in Section 3.

4. Grouping Hand Shapes

Training set of 3000 hand images such as those shown in Figure 2 and 3 were automatically grouped from the captured images (of video).To extract them, the Gaussian function is being used and by that the hand regions are extracted. In this project, the training images will be grouped together with the other images that shares same appearances by using the technique by [4], which is the binary images. Only after that, the k-mediod is then used to cluster hand images based on their shape.

4.1 Shape Context

Shape context is feature descriptor that named by [1] in their year 2000 paper. It is a way to describing shapes for object recognition. In shape context analysis, a set of points that lies on the known contour of unknown shapes is randomly selected. Then, the shape context of each point will be calculated before they are being matched. The shape distance between each

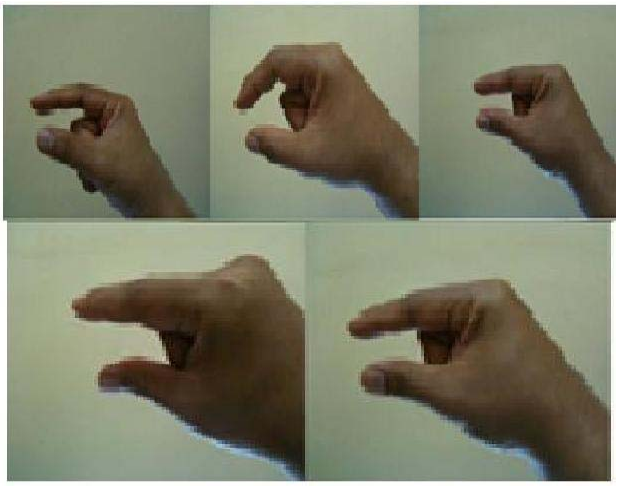

Figure 3. Hand gesture for alphabet H

point's pair is calculated and finally a nearest-neighbor classifier will be employed to indentify between the unknown shape and the known shape.

4.2 K-Medoid Clustering Algorithm

K-medoid and K-means are basically related. K-mean clustering is a method that partitions the observations into their identical (in terms of nearest mean) clusters. Both are partitional in minimizing squared error, distance between labeled points and a point that labeled as center of that cluster. In this paper, the result of clustered hand images can be seen in Figure 2 and 3. It is known that the grouped images are similar in each cluster.

5. Experiments

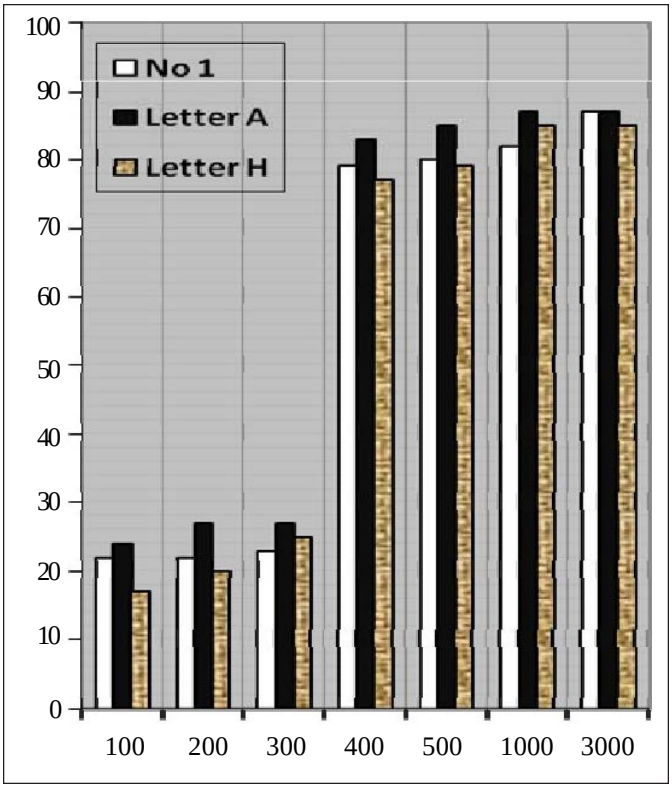

Figure 4. Experiments Result

From Figure 4, the *X*-axis defines rates of recognition and the *Y*-axis as the number of training images used. Blue bars(first from left) denotes the recognition of *Number* 1, Pink bars as *Letter A* and the last bars as *Letter H*. As the number of training images increases, the recognition rates increases too. However the high recognition rate are less significant upon using 500 training images until 3000 images.

5.2 Hand Detector and Shape Detector

From the experiments, total of 3000 images were captured from 2 different MSL signers that showing alphabet *H, A* and number 1.Out of that number, only 500 has been selected for training images. For the general hand detector, cascades of 9 layers were trained with total of 450 weak classifiers. It is found that the detection test rate was less than 1% which means that the data we captured were reliable. To train the branches of shape detector, the training and the test database were combined, together. Cascades of strong classifiers were trained on the images just right after each clustering.

From the Figure 4, our experiments are to detect and classify number 1, *letter A* and *letter H*. We also find out that if more than 500 training images used the increase rate of recognition slightly increase but however the calculation period is increased one fold. Thus we only use 500 training images instead of all 1500 images we captured earlier.

6. Conclusions and Future Work

By employing the skin color detection method in the first step, we believe that it will help in the future work of applying the whole concept for a larger sign language translator database as there are lots of hand movements and between each gesture, there will be frames that the hands will be absent from the image. By not employing this stage, the system will have to run to the middle of the process before discarding the image and thus the load on the machine would be higher than it should be; affecting other issue such as the response time.

The *k*-mediod clustering algorithm had definitely increased our recognition rate at affordable mechanisms.The usage of

boosted hand detectors' tree helped in the very first phase of the system. Instead of using only 9 layers and 450 weak classifiers, in the future we would hope to increase to more layers and classifiers so that the recognition rate can at least achieve 90% of success. However, to increase the classifiers in simple background is not an issue but what if the collected image in complex background that has clutter background, lighting intensities issue and also moving object? All these questions will be tried to be answered hopefully in the future research.

References

[1] Belongie, S., Malik, J., Puzicha, J. (2000). Shape context: A new descriptor for shape matching and object recognition. *In*: NIPS, p. 831–837.

[2] Friedman, J., Hastie, T., Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28 (2) 337–374.

[3] Y. Fruend and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55 (1) 119–139, 1997.

[4] Lockton, R., Fitzgibbon, A. (2002). Real-time gesture recognition using deterministic boosting. *In*: Proc. of British Machine Vision Conference.

[5] Schapire, R. (2002). The boosting approach to machine learning: An overview. *In*: Proc. of MSRI Workshop on Nonlinear Estimation and Classification.

[6] Schapire, R., Singer, Y. (1998). Improved boosting algorithms using confidence-rated predictions. *In*: Proc. of the 11^{th} Annual Conference on Computational Learning Theory, p. 80–91.

[7] Sherrah, J., Gong, S. (2001). Continuous global evidence-based bayesian modality fusion for simultaneous tracking of multiple objects. *In*: Proc. IEEE International Conference on Computer Vision.

[8] Starner, T., Pentland, A. (1995). Real-time american sign language recognition from video using hidden markov models. *In*: Proc. of SCV95.

[9] Viola, P., Jones, M. (2001). Robust real-time object detection. *In*: Proc. of IEEE Workshop on Statistical and Computational Theories of Vision.

[10] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *In*: Proc. of IEEE International Conference on Computer Vision, p. 734–741, 2003.

[11] Wren, C., Azarbayejani, A., Darrell, T. (1997). Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7) 780–785.

[12] Raja, S. M. Y., Gong, S. (1998). Tracking and segmenting people in varying lighting conditions using colour. *In*: Proc. IEEE International Conference on Automatic Face and Gesture Recognition.

[13] Zhang, Z., Li, M., Li, S., Zhang, H. (2002). Multi-view face detection with floatboost. *In*: Proc. of the Sixth IEEE Workshop on Applications of Computer Vision.

[14] Ong, E.-J., Bowden, R. (2004). A boosted classifier tree for hand shape detection. *In*: Proc. FGR.

[15] Abbas Cheddad, Joan Condell, Kevin Curran, Paul Mc Kevitt. (2009). A skin tone detection algorithm for an adaptive approach to steganography, Signal Processing, 89 (12), Special Section: Visual Information Analysis for Security, December , p. 2465-2478.