Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HAND GESTURE ACTIVITY BASED SIGN LANGUAGE RECOGNITION MODEL DEVELOPMENT USING SKIN PATCHES COLOUR DISTRIBUTION HISTOGRAM APPROACH

DR. DEEPAK ARORA[1], ASMITA DIXIT[2]

Computer Science and Engineering Department,
Amity University, Lucknow
E-mail: [1]deepakarorainbox@gmail.com, [2]asmitadixit5778@gmail.com

ABSTRACT

The advent of modern sciences has played a pivotal role in exploring other, if not better, avenues like the vision based hand interface techniques instead of the more traditional means of those of keyboard, mouse, pen etc, wherein the former has attained a higher relevance because of its natural and efficient outcomes. Through this paper an attempt has been made at developing an algorithm capable of intercepting the hand moments and locating its centroid. The proposed method tabulates the corresponding feature attributed to a gesture, after analyzing and intercepting the distance between the finger tips, their count, relative distance between the finger tips and their distance from the centroid, which eventually adds up to creation of the histogram of an image through the substantial and collateral information received through the hand gestures. The proposed technique works against a yellow background with interception of gestures of the front and the back side of hand. In an attempt of enhancing the robustness of the processing in consideration of the environmental impact through noise, the HSV format is taken into use. This method works on intercepting dedicated gestures of fingers to identify different alphanumeric characters followed by tracking of hand using motion and color cues. We have investigated the application of the method in sign language recognition by dedicated hand gestures feature learning. Through the instant method a proposal to an approach for recognition of sign language having usage in machine learning problem is being made which undoubtedly would be extremely helpful for deaf people to interact with others who are unaware of or don't understand Sign Language.

Keywords: *Machine Learning, American Sign Language (ASL), Skin Detection, Gesture Recognition, Image Tracking, Histogram.*

1. INTRODUCTION

The progress in human-machine has developed a very important user interface technology. One of the latest interface technologies is physical gestures because it makes the interaction process easier and due to this
human can more naturally control computers or other devices. Some of the applications are telerobotics where robots remotely follow the motions of master's hand [1]. Other example of hand gestures is character-recognition in 3-D space [2], [3], gesture recognition to control the functioning television [4], hand motion as a 3-D mouse [5].

There are two types of recognisation of hand gestures. One is gestures made from different arm positions and other is hand gesture made by different shapes from hand. Both of these gestures have many different applications in real world. We have focused on implementation of hand gesture recognition in real time with video processing for second type of hand gesture i.e. gesture from different shapes of hand is implemented with an application of Sign language. Gesture recognition systems identify human gestures and the information they convey. Use of gesture as the primary source of command input to computers has rapidly progressed in some areas such as computer simulation, by relying on special hardware and wearable devices [3, 4]. This method is often not cost-effective and is infeasible for some applications; consequently, gesture recognition based on substitute methods of data acquisition is being considered. In this paper the authors introduce a method for gesture recognition in 2D [5] space which authors used for interpreting hand and arm movements as gesture commands to a vision-based user interface.

Vision based hand gesture recognition has two categories: 3D model based methods and appearance model based methods. 3D model may exactly describe hand movement and its shape, but most of them are computational expensive to use. Recently there are some methods to obtain 3D model with 2D appearance model such as ISOSOM and PCA-ICA in [6] and [7].The authors have focused on appearance model based method.

Freeman et. al. recognized gestures for television control using normalized correlation [8]. This

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

technique is efficient but may be sensitive to different users, deformations of the pose and changes in scale, and background. Cui proposed a hand tracking and sign recognition method using appearance based method [9]. Although its accuracy was satisfactory, the performance was far from real-time. Just et al introduce modified census transform into hand gesture classification [10]. For the purpose of classifying each gesture respectively, their method obtains fairly good results. While the performance in recognition experiments were not so satisfactory and the recognition result of different gesture were of great disparity. Elastic graphs were applied to represent hands in different hand gestures in Triesch's work with local jets of Gabor filters [11]. It locates hands without separate segmentation mechanism and the classifier is learned from a small set of image samples, so the generalization is very limited.

The performance of vision based gesture interaction is prone to be influenced by illumination changes, complicated backgrounds, camera movement and specific user variance. Effective efforts have been made by many researchers to deal with these problems. Scale-space color features to recognize a hand gesture was used by Lars and Lindberg [12]. In their method, gesture recognition method is based on feature detection and user independent while the authors showed real-time application only under uniform backgrounds. Mathias and Turk developed a vision gesture interface with extended Adaboost for wearable computing, named HandVu [13]. It's insensitive to camera movement and user variance. The hand tracking acquired promising results, but the segmentation was not so reliable.

In this all hand gesture images are required to have the similar aspect ratios, which restrict the scope of applications.

2. PROPOSED WORK

Hand detection is of utmost importance and also the most critical for successful gesture recognition. In authors method the background being kept constant, allows the hand to be segmented as foreground. This method yields substantially decent results but often leads to the recognition of user's as foreground which is not as required. To bye-pass this limitation, HSV (Hue Saturation Value) color space is deployed which segments the skin color and recognizes the hand. HSV has better illumination resistance and segmentation capability in comparison with the RGB and YCrCb color spaces. Results are exceptional when the light source is behind the webcam but being from another direction (especially depicting from above), the result is poor. HSV color method is deployed to demonstrate the rest of the work. Thereafter the segmented hand is converted to a binary image. For an efficient tracking, it is advocable not to have the other hand and user's face in the image being captured through the webcam. Thereafter gestures formed with different hand shapes are recognized and after implementing this in real time the authors have deployed the same for the sign language implementation. A camera was used to take images and MATLAB being used as programming toolbox. In this relation the hand in the frame of the video thereby generated a radial histogram for that detected hand. The techniques for final recognition used herein are features based on histogram.

3. FEATURE EXTRACTION

A histogram has been generated dedicated for recognizing sign language by hand gestures. To achieve this end the distance as well as angle of each boundary point from hand's centroid are calibrated and calculated. Then, making a matrix of size 361 where indices of matrix correspond to different angles and each index of matrix contains corresponding distance. Sometimes, for a particular angle, the distance may vary as there may be more than one distance. In that instant case, maximum distance is taken into consideration. Band-centroid to Hand-centroid angle is calculated. For the purpose of calibrating hand for different hand yaw 0 has been used as reference angle shifting the whole radial histogram by hand-centroid to band-centroid angle.

In some of the literatures [14] Haar-like features and the AdaBoosting algorithm, a set of hand gestures have been used for recognizing a real-time image. That has converted to gesture commands to control and manipulate the digital signage display.

In further section results relating to the algorithm have been shown finally comparing the results to Haar-like feature based AdaBoosting algorithm. In another literature [15] Thinning algorithm is used which a morphological operation is using which a binary image is thinned by repeatedly deleting pixels inside the shape.

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(a) Haar-Like Feature Extraction with AdaBoosting

To solve the problems faced by color and shape based algorithms, a set of Haar-like features is employed which have been successfully used for face detection. Each Haar-like feature is a template of multiple connected rectangles of black and white. The value of a Haar-like feature reduces to be the difference between the sums of the pixels' values within the black and white rectangular regions:

$f(x) = W_{black} \cdot \sum$ (pixel value) - $W_{white} \cdot \sum$ (pixel value)

black region white region

Where W_{black} and W_{white} are categorised as weights meeting the compensation condition:

W_{black} .black region = $W_{white.}$ white region

A single Haar-like feature is certainly not enough in identifying the object with a higher rate of accuracy.

The technique of boosting is employed to supervise the machine learning algorithm to improve the overall accuracy stage by stage based on a series of weak classifiers. A weak classifier being trained with a set of training samples at each step.

Thereafter the instant trained weak classifier is then added to the final classifier with a strength parameter in proportion to the accuracy of the weak classifier. The training samples missed by the current weak classifier are re-weighted with a bigger value and the future weak classifier will attempt to fix the errors made by the current weak classifier improving the overall accuracy.

Adaboost algorithm thereafter picks the Haar-like feature that yielding the best classification accuracy in the first iteration. The weak classifier based on this Haar-like feature is added to the linear combination with a parameter in proportion to its accuracy.

In the second iteration, the training samples are re-weighted: training samples missed by the first weak classifier are boosted in importance thus causing the second Haar-like feature to pay more attention to these misclassified samples. To be selected, the second Haar-like feature must achieve a better accuracy for these misclassified training samples achieving the reduction of the overall error. This iteration goes on by adding new weak classifiers to the linear combination until the

Terminally required accuracy is met on the whole. The final training result is a strong classifier composed by a linear combination of the selected weak classifiers.

(b) Thinning algorithm based sign language recognition.

In this thinning of images is done to remove from binary images, selected foreground images in order to obtain a skeletal image by compressing all lines to single pixel thickness. In this, the terminating points are determined by the end points of the fingers. For pattern recognition and character recognition Hither to thinning algorithm has been used.

Thinning is performed by iteratively removing pixels inside the shape to reduce it without shortening or breaking apart.

(c) Histogram Based Feature Extraction

The authors approach as discussed in previous section works on determining number of fingers deploying boundary imaging technique for a particular hand gesture. For finding the boundary several image process operations are performed. First of all finding of RGB range in a RGB cude of the front/back hand and band are recorded from images focused at the centroid. After that the algorithm removes the outer area means edge area of both hand and band image from colour cude values of these particular regions. After this B/W image framing is also performed.

This refined and binearized hand boundary image is thereafter processed through image morphological operations. These operations are morphological opening on the greyscale or binary image with the element of structuring. The morphological open operation is dilation preceded by erosion, deploying the same structuring element for both operations. The structuring element used here is disk type bearing block size 3. Another operation is closing on the gray scale or binary image, returning the closed image. The structuring element must therefore be a single structuring element object, as supposed to an array of objects. The morphological close operation is erosion preceded by dilation, employing the same structuring element for both operations.

Journal of Theoretical and Applied Information Technology
10th November 2014. Vol. 69 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The structuring element used here is also disk type bearing block of size 3.

After that creation of Histogram has specified and calculated the peaks in the image. Furthermore the algorithm deduces and calculates the angle and distance of each peak.

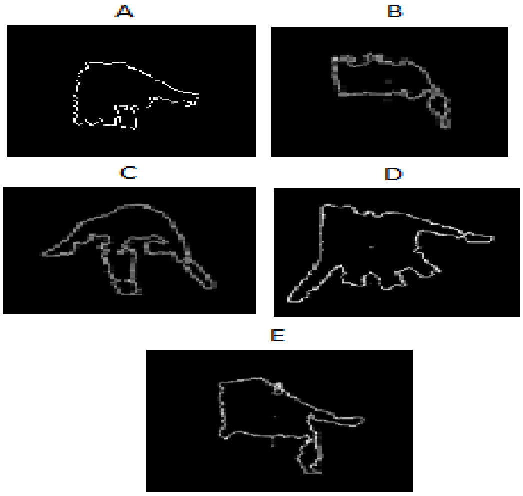
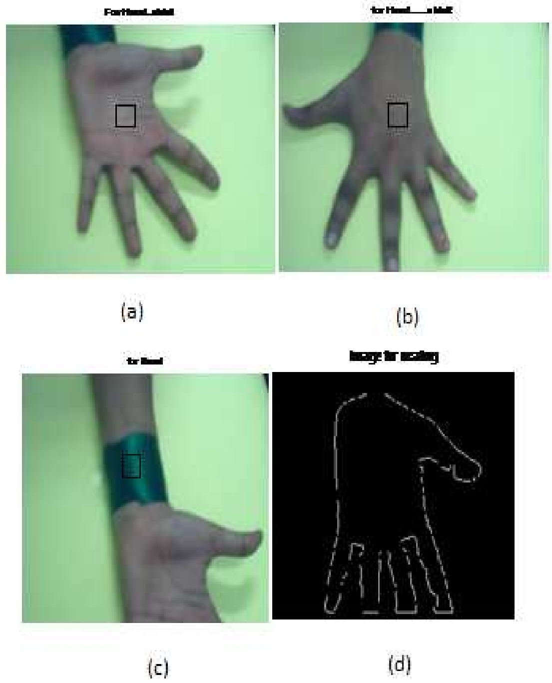


Figure 1: (A) Front Hand (B) Back Hand (C) Band With Centroid And (D) Hand Boundary Extraction.

4. SIMULATION RESULTS

By an RGB to HSV format, employed primarily to convert image of both sides of hand and the image of green colored wrist band to identify the shape of input handoff the user. Care has been taken that the background should be yellow and the centroid of hand should match the centroid of capture frame see figure 1.Figure 1(a) exhibits the front hand shape where centroid of hand coincides with the square box placed at the centre of camera image frame. Similarly the authors have displayed figure for back side and band image with centroid focused position. Figure 1(d) exhibits the hand after tracking the RGB cude index of hand and band color. This image binarises the pixels to one which belongs to the RGB cude of hand reducing other pixels to zero. For achieving the hand envelope of figure 1(d) centroid of hand and band are calibrated and calculated for evaluating the distance between finger tips and there distance from the centroid.



Thereafter for the matching stage the same procedure has been followed again for five different kinds of hand gestures (Figure 2).

Figure 2.Five Different Gestures That Are Used To Recognize Different Signs Language.

Figure 2 shows the five gestures that the authors have deployed to represent the five signs. Each gesture is recognised in particular through an algorithm.

Figure 2(a) represents the thumb for this gesture our code shows the character A. Figure 2(b) the sign is a large finger being the sign of B. Figure 2(c)
represents C and thereafter, respectively for figure 2(d) sign is D, for figure 2(e) sign is E.

The authors have designed an AVR microcontroller based hardware that can be interfaced using UART port with the laptop for the purpose of establishing communication with serial port signals. This system accesses a definite binary code for aforementioned five gestures. Using this code five control signs on the LCD display connected to microcontroller output port can be displayed.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

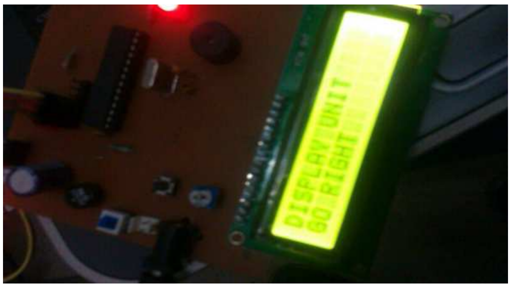


Figure 3.AVR Controller Based LCD Display Functioning On The Hand Gesture Signs.

Figure 3 shows the AVR controller based circuitry that produces five different control signals which that can be created for communicating with deaf people. The five communication signal for our hand gestures are A-Go Straight, B-GO Left, C-Stop, D-Go Right and E-Go Left that are respectively displayed on the LCD screen.

5. EFFICIENCY COMPUTATION

The efficiency of the proposed and implemented prototype model is calculated for five different video inputs of alphabets (A to Z) of the American Sign Language.

An expected result (ER) is evaluated when a sign is recognised correctly and a wrong result (WR) is evaluated when the sign is unidentifiable or not recognised properly.

Formula for efficiency calculation is:
ER/(ER+WR)*100

Table.1. Efficiency Calculation For Alphabets

Video Trial	**ER**	**WR**	**ER/(ER+WR)**	**(%) Efficiency**
Video 1	10	0	1.0	100
Video 2	10	0	1.0	100
Video 3	10	0	1.0	100
Video 4	8	2	0.8	80
Video 5	10	0	1.0	100

6. RESULTS AND DISSCUSSION

Drawing a comparison, between the sign language recognition which uses the method of thinning algorithm, wherein the approach adopted is primarily based on the concept of skin colour extraction whereby the required hand image can be invariably extracted from the background noise whereas the thinning approach more so, is largely concentrated on the application of the morphological operation of iterative deletion of the pixels inside the shape to achieve a shrinking result bypassing the shortening or breaking apart, which increases the possibility of the distortion in the recognition of character or numeral. Moreover the Haar-like figure attributed to only 50% of the efficiency until in collaboration with the Adaboost Algorithm thereby resulting in the efficiency rate along almost 97%.

The Authors approach uses, colour cubes of the skin and band colour region for histogram plotting. The concentration, for an efficient result, is maintained to a number of contingencies like the number of fingers count, the distance calculation from hand to band region and centroid calculation. Further the RGB to HSV conversion is employed resulting the skin coloured region to be easily captured thereby resulting in the removal of noise whereas the above two approaches do not utilise the skin colour based feature extraction. Furthermore the shape dimension used is disc, on the contrary thinning largely depends on the usage of square and line shaped dimensions whereas the Haar-like gesture changes to rectangle thereby creating tables respectively of the edge features. In comparison to the efficiency which is somewhere around 95-99% in student constraints, the same is a meagre 50% of a Haar-like feature invariably giving weight to the argument that the Authors algorithm is far superior to Haar feature. Nevertheless, equivalent performance been dispensed by the Authors Algorithm when in comparison with the Haar Like Adaboost Detection Algorithm which possesses an efficiency rate of 97-99%. Moreover the Authors work provides equivalent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

efficiency of about 95-96% similar to that of the Thinning Algorithm.

7. CONCLUSION

In conclusion an algorithm has been created with the capacity of sign recognition based on hand gestures which can impart information with AVR microcontroller representing five different activities being displayed on LCD. The algorithm employs radial histogram approach to calculate distance between finger tips and also there angle from each other and the centroid of the hand Using number of fingers and there distance between finger tips wherein the algorithm generates five different characters and control signals for the LCD display. The algorithm is robust and prone to noise and does not require any dedicated training data prior to initiation of different sign languages. Our results have been tested for different lightning conditions and different set of hands. The efficiency ends up being somewhere around 95-99% under student constraints.

The authors have been successful in adducing a simpler and fast algorithm with equivalent efficiency for controlling signals of LCD display by hand gesture as to a Haar like Adaboost Detection which requires a large training database. Also in comparison to thinning algorithm approach the efficiency is 95-99%. The output obtained is satisfactory and the work is implemented successfully without complexities in better ways.

REFERENCES

[1] T. H. Speeter, “Transformation human hand motion for telemanipulation,”Presence, vol. 1, no. 1, pp. 63–79, 1992.
[2] S. Zhou, Z. Dong, W. J. Li, and C. P. Kwong, “Hand-written character recognition using MEMS motion sensing technology,” in Proc.IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics, 2008, pp.1418–1423.
[3] J. K. Oh, S. J. Cho, and W. C. Bang et al., “Inertial sensor based recognition of 3-D character gestures with an ensemble of classifiers,” presented at the 9th Int. Workshop on Frontiers in Handwriting Recognition, 2004.
[4] W. T. Freeman and C. D. Weismann, “TV control by hand gestures,” presented at the IEEE Int. Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, 1995.
[5] L. Bretzner and T. Lindeberg, “Relative orientation from extended sequences of sparse point and line correspondences using the affine trifocal tensor,” in Proc. 5th Eur. Conf. Computer Vision, Berlin, Germany, June 1998, vol. 1406, Lecture Notes in Computer Science, pp.141–157, Springer Verlag.
[6] Haiying Guan, Rogerio S. Feris, and Matthew Turk, “The isometric self-organizing map for 3d hand pose estimation,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 263–268.
[7] Makoto Kato, Yen-Wei Chen, and Gang Xu, “Articulated hand tracking by PCA-ICA approach,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 329 – 334.
[8] W. T. Freeman and C. Weismann, “Television control by hand gestures,” in Proceedings of International Workshop on Automatic Face and Gesture Recognition. Zurich, Switzerland, June 1995, pp. 197–183.
[9] Y. Cui and J. Weng, “View-based hand segmentation and hand sequence recognition with complex backgrounds,” in Proceedings of 13th ICPR. Vienna, Austria, Aug. 1996, vol. 3, pp. 617–621.
[10] Just A., Rodriguez Y., and Marcel S., “Hand posture classification and recognition using the modified census transform,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 351–356.
[11] J. Triesch and C. von der Malsburg, “Robust classification of hand posture against complex background,” in Proceedings of Int. Conf. on Face and Gesture Recognition. Killington, Vermont, Apr. 1996, pp. 170–175.
[12] Lars Bretzner, Ivan Laptev, and Tony Lindeberg, “Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering,” in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Washington D.C., May 2002, pp. 423–428.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[13] Mathias Kolsch, "Vision based hand gesture interfaces for wearable computing and virtual environments," PHD Dissertation, UCSB, 2005.

[14] Qing Chen, Fran¸ cois Malric, Yi Zhang, Muhammad, and "Interacting with Digital Signage Using Hand Gestures." Springer2009volume. [15] S.N. Omkar, M. Monisha, "Sign Language Recognition using Thinning Algorithm." ICTACT journal on image and video processing, Aug. 2011, vol.2