ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

Wielding Audio-Books for Visually Impaired using Gesture Recognition

Atul Dhingra[1], Kumar Vishal[2]

Student, Netaji Subhas Institute of Technology, Delhi, India [1]

MS Student, International Institute of Information Technology, Hyderabad, India [2]

Abstract: The paper presents various aspects of hand based gesture recognition. The paper discusses two main modules, colour-based tracking and colour-independent tracking. Various aspects of both these modules are discussed in detail in this paper. In the case of Colour Based Tracking we have used RGB model to segment out hand to track gesture. In the latter module, we use a combination of Mixture of Gaussians Background (MOG) model and convex hull followed by convexity defects to segment the hand. K-means is used to track the centre of the region of interest (ROI).

Keywords: Convex Hull, Convexity Defects, K-means, RGB Model

I. INTRODUCTION

A sequence of hand gesture is one of the most inherent features of human-human interaction. This particularly forms a backbone for the speech and hearing impaired people using American Sign Language (ASL) to interact with the outer world. There has been therefore a lot of interest to emulate the same model in human-computer interaction to provide a more intuitive user experience to this section of society. Also, in the realms of virtual reality, gesture technology has seen a wide rise of applications, providing easy accessibility to the users. But, it is also widely believed that Human Computer Interaction (HCI) may become the bottleneck of all the flow of the existing knowledge [2]. The previously known HCI devices like keyboard and mouse have become very familiar but lack the requisite speed and interface to realize the level of research in the realms of virtual reality. Although there are a lot of efforts going on in these areas, we focus our attention primarily on developing our research in aim to ameliorate the accessibility to visually impaired people.

Both the modules discussed in this paper have their own significance and shortcomings, and it entirely depends on the type of application on the choice of module to be used. Although coloured model of tracking is more accurate because it can be segmented primarily on the basis of colour models, but it is restrained in the sense that we need to have the colour marker at all the times to make the gestures which beats the purpose of in our application, as it restricts the accessibility to the visually impaired. The model independent of colour though makes up for the shortcomings of the colour based systems but it is less robust and responsive than the colour based model. This is due to the fact that it is not possible to accurately segment out the hand in uncontrolled and complex [1] background. In this paper we have used a combination of Background MOG model with Convexity defects and k-means algorithm to track the hand in real time. This is discussed in detail in section 4.The rest of paper is organized as follows. Section 2 talks in detail about the previous works and motivations for the research. Gesture recognition techniques are explained in detail in section 3. Section 4 describes the experiment in detail. Conclusions and future works are presented in section 5.

II. PREVIOUS WORKS

One of the most common methods of classification of gesture recognition methods is on the basis of input [5]. This can be achieved either by use of gloves and sensors or by use of computer vision algorithm. The use of gloves provides a better accuracy of the gesture but is restraint on freedom of motion as it is connected to various electrical components for transmitting the signals from the sensors to the computer. Such type of systems caters only a highly specific type of application where accuracy is of utmost importance as in the case of medical surgery in virtual environments. The computer vision algorithms, though not as robust as former method but allows a greater ease of

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

access to user as it includes no fixed wire components. A camera mounted on a standard laptop is used to locate the region of activity and extract the ROI using Computer Vision algorithms. Simple as it may sound, but to accurately segment out the hand is a very specific process and research is still going on in this direction to bolster the level of accuracy for segmenting the hand properly. Although some of the best results have been observed with the use of kinect sensors [3] and therefore most popularly used in virtual reality gaming [4], this would ensue a separate hardware arrangement and defeat the aim of our research to make use of the VGA camera input from a standard laptop, which is readily available hardware. The most recent foray into the gesture recognition system providing accurate gestures using a camera from laptop has been by FlutterApp [8], which can work on as low as a VGA camera inbuilt to laptop to sense gestures. Although it has a limited number of gestures and works on static gestures as compared to dynamic gestures as in our case, the accuracy of the system is highly commendable. This has thus motivated us towards this direction of research.

III. METHODOLOGY

Hand Gestures can be detected using various methods [3] depending upon the heuristics. Some the most popular methods are explained in this section.

3.1 Detection

Detection forms the first step in the case of gesture recognition. In our research, our area of interest is hand, which needs to be detected with utmost accuracy. This is the most important step in our project, as all the consequent processes depend on how accurately the hand has been detected.

3.1.1 Colour Based

Skin colour segmentation has been one of the most widely used methods to segment out the hand from the entire frame which has a back-ground devoid of the same colour. Other colour spaces that have been used for similar process are HSL, HSV, yCrCb, normalized RGB. For example if we want to segment out a red-coloured marker, we can perform it on the basis of HSV model to segment out the red colour marker. Hue and saturation values are adjusted such that it corresponds to a red coloured object. Generally a coloured marker is used if we are performing gesture recognition using hand, as in most of the frames face will also be present and hence hand cannot be segmented out with high accuracy. These methods are used in the case where a very accurate gesture is to be achieved, but these methods are constrained by the fact that it requires a coloured marker to work with at each point of time and it number of restrictions seem to accrue with such an approach.

3.1.2 Shape Based

The characteristic shape of the hand can be utilized to detect them in images, and contours can be extracted thereafter. Once the contours are extracted on the basis of shape of hand the only task remains is to detect the centre of the resulting contour in real time motion or to analyse it frame for frame to observe the direction of motion of the centre and hence the gesture. This reduces the complex problem of tracking a hand to tracking just the centroid of the contour. Active contours or convex hull with convexity defects can also be used for better segmentation of hand.

3.1.3 Motion Based

Motion based detection is one of the more difficult to achieve as it has many constraints on the motion, the major constraint being that only hand is to remain in motion in the entire frame and the background remains the constant. This is particularly hard to achieve in an uncontrolled environment [1].

Several other methods have been employed including 3D hand based models, and learning detectors from pixels. The latter is achieved by training hand gestures by training classifiers. The basic assumption is that hand appearance differs more among hand gestures than it differs among different people performing the same gesture [6]. More recently, methods based on machine learning have started to be used for the purpose the most popular methods being Boosting.

3.2 Tracking

The next step towards hand gesture is tracking the segmented hand frame for frame. This is a vital step which provides the result in the form of the type of gesture performed by the user. These gestures may be used in a raw form or to track a specific gesture in high end applications of gestures like as in Air Touch [7]. In this paper we make use

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

of the raw gesture, as our basic aim is to navigate through audio-books, which can be achieved by much simpler and crude gestures. We have a few methods that are most widely used in tracking the moving gesture, some of which are explained as follows.

3.2.1 Template Based Matching

Template based matching is a technique that is used in classifying objects. Template matching techniques compare portions of images against one another and make decisions in the form of correlation. In this method, sample images are used to recognize similar objects in source image. If standard deviation of the template image compared to the source image is small enough, template matching may be used. The matching process moves the template image to all possible positions in a larger source image and computes a numerical index that indicates how well the template matches the image in that position and match is done on a pixel-by-pixel level. When using template-matching scheme on grey-level image it is unreasonable to expect a perfect match of the grey levels. Instead of yes/no match at each pixel, the difference in level should be used. Correlation is a measure of the degree to which two variables agree, not necessary in actual value but in general behaviour.

Figure 1: Shows the template and corresponding input image

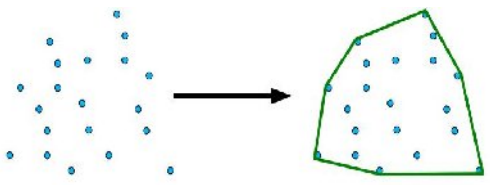
3.2.2 Mean Shift Algorithm

Mean shift is a procedure for locating the maxima of a density function given discrete data sampled from that function. It is useful for detecting the modes of this density. The mean shift algorithm can be used for visual tracking. The simplest such algorithm would create a confidence map in the new image based on the colour histogram of the object in the previous image, and use mean shift to find the peak of a confidence map near the object's old position. The confidence map is a probability density function on the new image, assigning each pixel of the new image a probability, which is the probability of the pixel colour occurring in the object in the previous image. A few algorithms, such as ensemble tracking, CAMshift, expand on this idea. In this paper we have made use of k-Means clustering algorithm [9]. K-means clustering is a method of vector quantization originally from signal processing that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The problem is computationally difficult, however there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centres to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

3.2.3 Contours

In our paper we have used convexity defect points to work with. To find the defects we need to calculate the contour and hull with respect to which defects are calculated. These are described in short.

A contour is a list of points that represent, in one way or another, a curve in an image. This representation can be different depending on the circumstance at hand. There are many ways to represent a curve. Contours are represented by sequences in which every entry in the sequence encodes information about the location of the next point on the curve. The convex hull of a set Q of points is the smallest convex polygon P for which each point Q is either on the boundary of P or in its interior[10] Hull.png



ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

Figure 2: A convex hull

Convexity Defects is a feature that find the defects between a convex hull and a contour; those defects are useful to find features in a hand, as for example the number of fingers.

Figure 3: Convexity Defects

These convexity defect points are cardinal to our aim in the project as these are points on the basis of which the decision of k-means centroid is taken. We prefer defect points over the convex hull points because the defects are a better measure to input into k-means. Suppose we have an outlier far away from ROI, and we use k-means to cluster the points, the centroid of both the clusters will lie outside ROI as the hull from noise to the ROI is comparable to that covering ROI alone. But, in the case of convexity defect points, we will obtain defects on hand and an outlier far away from hand which will cluster the noise away from ROI and the centroid of cluster on ROI will be on the hand, which can be used to track the gesture.

3.3 Pre-processing

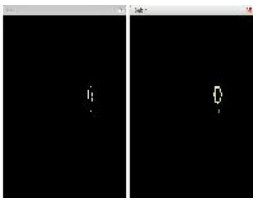
In this paper we have made use of the convexity defect points and used them for clustering into two, using k-means algorithm, breaking it into ROI and other noisy data points. We then use the centroid of ROI to track the gesture. Simple as it may seem, this process is highly affected by noise and hence pre-processing of noisy data is of cardinal importance. In this experiment, k-means is in-fact used as a measure to reduce noisy data points, which is coupled with hard thresholds to restrict any noisy data point to be added to frame buffer, working of which is explained in detail in section 4.2. Pre-processing of data points is thus an important part of the pro-cess which is directly linked to the accuracy of the gesture

IV. EXPERIMENT

The experiment is divided primarily into two major parts: a coloured based model and a model independent of colour. Both these methods are talked about in detail in the section below

4.1 Colour Based Module

One of the most robust methods to track hand gestures is to track a coloured marker placed on the hand in motion. The coloured marker may be of any colour which can be exclusively represented under various lighting conditions with a wide range of Hue, Saturation and Value. Generally this colour is taken one of the primary colours of RGB spectrum, and more often than not, the choice is red. In our experiment, we have used easily available Light Red markers made of Stick On pads and RGB values are in turn calculated to represent the model very accurately so that the colour can be segmented out effectively. The colour based gestures are more sensitive and much more gestures can be obtained in this case as compared to colour independent gesture module. In this experiment we use RGB colour model to segment the red coloured marker. The red colour component obtained is subtracted from the grayscale image of the same input so as to segment out the red coloured marker. The segmented ROI is used in turn to find the contours. We obtain the centre of the contours and track it in real time to find the resultant gestures. We perform the operations frame for frame and keep a specific window which is determined heuristically to check whether a legitimate gesture is performed or it was due to some random noise. We choose a window such that small random motion is not taken into consideration and hence obtain the gestures for maximum number of trials. Some error might creep up if the back-ground has other red objects, which will also be detected. If non-ROI portion of the image also has red component then it will nix the algorithm employed. For such cases we have worked on background model of MOG. A number of trials were performed, out of which accurate gesture was recorded 95% of the time.



ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

(a) (b)
Figure 4: (a) Input Image, (b) Output contour and bounding box

4.2 Colour-Independent Module

A colour-independent module based on vision algorithm and making it robust at the same time puts forth a challenge and leaves us with limited number of options to work with. In such cases, prior knowledge of hand becomes a very important factor to perform the operation. But, as we are considering real time gestures and not static gestures like in the case of "FlutterApp"[8] we cannot rely on this method as we have to track the motion of the hand and not just the learn the different type of static hand. We start by training the background using Gaussian-Mixture based foreground/background segmentation algorithm with the help of OpenCV libraries. The contours are plotted on the learnt background model and the new input frame from the camera. Any change from the learnt background is detected by contours. To make the procedure more robust we keep a threshold on the area the contour takes. This eliminates some erroneous capture of points as contours. Now these contours are used to form a convex hull joining the outermost contour. Convexity defects are then calculated which provides us with the points of inflection at the convex hull. K-means algorithm is thereafter applied to cluster the points into two clusters, and obtaining the centre of the cluster which has majority of points, and hence removing the outliers. This reduces a lot of noisy data points that creeps up due to slightly moving background. Figure 5(c) shows the various defects points marked by red dots, and k-mean centre is shown by a yellow dot.

(a) (b) (c)
Figure 5: The figure shows (a) Foreground mask when no gesture is made, (b) Foreground image when gesture initiated, and (c) k-means centre of the useful defects

This centroid from the k-means is cardinal to the entire experiment. We track this point in response to the gestures that are made in real time. We have worked on 3 gestures overall, performing 4 different operations; Left to Right, Right to Left and Play/Pause gesture. To decide what gesture is made, we work on the values of co-ordinates provided by the k-means centroid. If the value is more than a specific window in x direction, we treat them as Horizontal gestures, depending upon the sign change of values. This operation is similarly performed in the case of vertical direction as well. A hard threshold is also kept on how the frame buffer is filled in gestures from x and y direction. If the k-mean point is observed to be at a specific point in one frame and it crosses a particular threshold (in this case 200 px), the point is discarded and treated as a noise. A number of trials were carried out and an accurate gesture was reported 90% of the time.

V. CONCLUSIONS

Two modules for gesture recognition were presented in this paper of whose merits and demerits were discussed in detail in accordance to the type of application. The gesture recognition application was converted to an API for an already existing application of Audio-book library at Center for Visual Information Technology at IIIT-Hyderabad, India. Though in our application, colour-independent module is more desired, but it still requires some more work before it can be fully realised to such a system with optimum level of accuracy. In the future we would like to work to integrate our method with some of the existing methods of template matching like Viola Jones Algorithm [11] that has shown great promise in the realms of face detection.

REFERENCES

[1] J. Triesch, C. von der Malsburg, “A system for person-independent hand posture recognition against complex backgrounds” IEEE Trans. PAMI 23 (12) (2001) 1449–1453

[2] V. Pavlovic, R. Sharma and T.S. Huang, “Visual interpretation of hand gestures for human-computer interaction: A review,” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol. 7(19), pp. 677–695, 1997

[3] Zhou Ren, Jingjing Meng, Junsong Yuan, Zhengyou Zhang, "Robust hand gesture recognition using kinect sensor" MM’11 Proceedings of the 19th ACM International conference on Multimedia, pp. 759-760

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. II, Issue V, May 2015

[4] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, Yael Edan, "Vision-based hand-gesture applications" Communications of the ACM, Vol. 54 No. 2, Pages 60-71.

[5] Claudia Nolker, Helge Ritter, "Detection of Fingertips in Human Hand Movement Sequences" Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, pp. 209-218

[6] X. Zabulis, H. Baltzakis and A. Argyros. "Vision-based Hand Gesture Recognition for Human-Computer Interaction." Computer Vision Techniques for Hand Gesture Recognition

[7] Daniel R. Schlegel, Albert Y. C. Chen, Caiming Xiong, Jeffrey A. Delmerico, Jason J. Corso, "AirTouch: Interacting With Computer Systems at a Distance", 2010 IEEE

[8] https:// flutterapp.com/

[9] J.A Hartigan and M.A Wong, "Algorithm AS 136: A K-Means Clustering Algorithm", Journal of Royal Statistical Society, vol 28, No.1(1979), pp. 100-108

[10] Cormen et. Al, "Introduction to Algorithms"

[11] P. Viola and M. Jones, "Robust Real-Time Face Detection," Int',l J. Computer Vision, vol. 57, no. 2, pp. 137-154, May 2004.