Available Online at www.ijecse.org ISSN- 2277-1956

Dynamic Vision Sensor Camera Based Bare Hand Gesture Recognition

Kashmera Khedkkar Safaya [1] *, Asst.Prof. Rekha Lathi* [2]

[1 2] *Department of Information Technology and Computer Engineering*

[1] *Pillai's Institute of Information Technology Research and Studies, India*

[2] *Pillai's Institute of Information Technology Research and Studies, India*

[1] *kashmera.k@gmail.com*

Abstract- This paper proposes a method to recognize bare hand gestures using a dynamic vision sensor (DVS) camera. DVS cameras only respond to pixels with temporal luminance differences, which can greatly reduce the computational cost of comparing consecutive frames to track moving objects. This paper attempts to classify three different hand gestures. We propose novel methods to detect the delivery point, to extract hand regions, and to extract useful features for machine learning based classification. In order to do this, the paper begins by trying to understand the importance of gestures and how humans use gestures to communicate.

Keywords – Dynamic vision sensor camera, Hand gesture recognition

I. INTRODUCTION

Computer is used by many people either at their work or in their spare-time. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse. Every new device performs more complicated communication with the computer. This has been possible due to the result oriented efforts made by computer professionals for creating successful human computer interfaces. As the complexities of human needs have turned into many folds and continues to grow so, the need for Complex programming ability and intuitiveness are critical attributes of computer programmers to survive in a competitive environment. The computer programmers have been incredibly successful in easing the communication between man and computer. In other areas where 3D information is required, such as computer games, robotics and design, other mechanical devices such as roller balls, joysticks and data gloves are used. Humans communicate mainly by vision and sound, therefore, a man machine interface would be more intuitive if it made greater use of vision and audio recognition. Another advantage is that the user not only can communicate from a distance, but need have no physical contact with the computer. However, unlike audio commands, a visual system would be preferable in noisy environments or in situations where sound would cause a disturbance. The visual system chosen was the recognition of hand gestures. The amount of computation required to process hand Gestures is much greater than that of the mechanical devices.

The rest of the paper is organized as follows. Proposed system is explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. PROPOSED SYSTEM

A. Hand gesture recognition –

The goal is to recognize three different bare hand gestures and use those gestures in mouse free interface (as right click, left click and mouse movement). These three different hand poses are Rock, Paper and Scissors Shown in fig

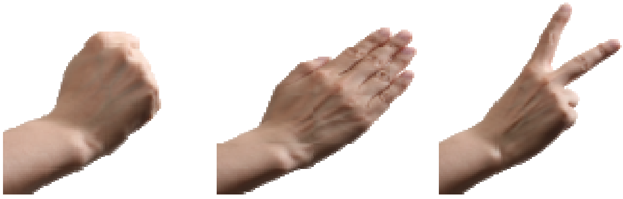


Figure1. Three Possible Throws: Rock, Paper, and Scissors [5]

To recognize a throw, the point where the player delivers a throw should be detected first; we call this point as delivery point. In terms of vision terminology, detecting the delivery point can be thought of as detecting a useful hand posture from hand gesture. Whereas hand posture refers to a static hand pose without involvement of movement, hand gesture refers to a sequence of hand postures involving dynamic movement [2]. Once the delivery point is detected, we can classify different throws using only a single posture captured at the delivery point. Previous literatures on using frame-based cameras discuss two main approaches for bare hand gesture recognition: model-based and appearance-based. The approach is similar to the appearance-based approach in a sense that a hand posture is classified based on the hand shapes. After recognizing the gestures as rock, paper and scissors theses gestures also recognized as movement of mouse, right click and left click respectively. In this when throw will be paper gesture it will recognize as right click and when it is scissors it will recognize as left click. The aim is to design mouse free interface and identify the hand poses or gestures. All stages of proposed systems are shown in following fig 2.

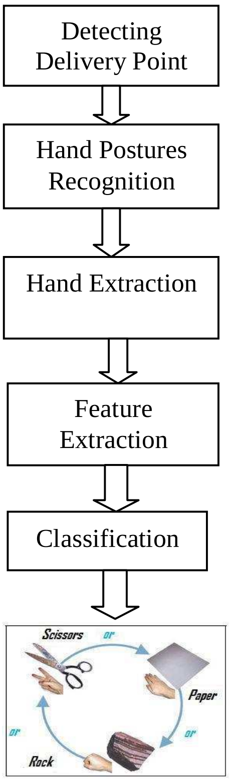
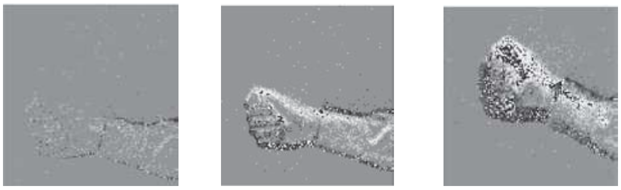


Figure2. Block Diagram of Proposed system

1.Detecting Delivery Point

The delivery point refers to a point where it delivers a throw. The movement of the hand is slower as the hand reaches the end of the delivery point in order to stop moving at the delivering phase. DVS cameras are only response to the pixels with luminance intensity, and send events of only those pixels. Since there is little movement of the hand at the delivery point, the number of events within a frame will be dramatically reduced. Once the delivery point is detected, system can classify different throws using a single posture captured at the time of delivery point. To find the delivery point, track the number of events for each frame. If the number of events in the frame less than the given threshold TH then the event is regarded as a delivery point. If TH is too large, the frame detected as delivery point can be different from actual hand throw. Even if the hand pose in the detected frame is the same as the actual throw, the frame may not be appropriate one for detecting hand pose. The reasons are that the frame is likely to contain too many events, which will increase the computational cost, and the shape of the hand can be not clear by thickening the boundary as shown in following fig 3(c). In contrast, if the size of TH is too small, as shown in fig

3(a) the frame detected as a delivery point may contain too few events to recognize the shape of hand. Therefore, TH needs to be carefully chosen. Fig3 (a) (b) and (c) shows the frames detected as a delivery point.



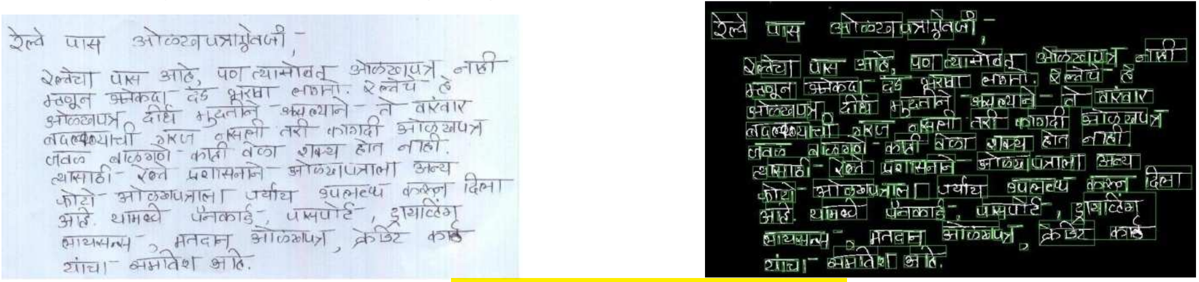
a) TH =1,000 b) TH=5,000 c) TH=10,000

Figure3. The Frames at the Delivery Points [1]

However stopping the hand movement does not mean that throw is delivered. Therefore it track the direction of movement for each frame, and the first frame with number of events less than TH during the downward movement is considered as delivery point. To summarize, the frame is regarded as a delivery point when it has the number of events less than the given threshold (TH) and the hand is moving downward, where the moving direction of the hand is determined as delivery point.

2.Hand Posture Recognition

Once the delivery point is detected, the frame at the delivery point is used to recognize hand pose. Since there are some noisy events, a noise filtering process is first conducted using connected component analysis. It is often useful to extract regions which are not separated by a boundary [3]. Any set of pixels which is not separated by a boundary is call connected. To apply connected component analysis, the stream of events of a frame is represented by a 128 by 128 matrix. There are two types of events: on and off events. Although a single matrix can be generated by ignoring the event types each of which represents on and off events are used, and connected component analysis is applied to each matrix separately. Since events of the same type are likely to occur closely in space, we expect that the two matrix representation can filter out noisy events that are surrounded by events of a different type. Connected component analysis is shown in following in fig-4.



रेल्वे पास ओळखपत्राबुलजी,

रेल्वेचा पास आहे, पण त्यासोबत ओळखपत्र नाही
म्हणून अनेकदा दंड भरावा लागतो. रेल्वेचे हे
ओळखपत्र दीर्घ मुदतीने असल्याने ते वारंवार
बदलण्याची गरज नसली तरी कागदी ओळखपत्र
जवळ बाळगणे काही वेळा शक्य होत नाही.
त्यासाठी रेल्वे प्रशासनाने ओळखपत्राला अन्य
फोटो ओळखपत्राला पर्याय उपलब्ध करून दिला
आहे. यामध्ये पॅनकार्ड, पासपोर्ट, ड्रायव्हिंग
लायसन्स, मतदान ओळखपत्र, क्रेडिट कार्ड
यांचा समावेश आहे.

Figure4. Connected component method to reduce noise

3.Hand Extraction

After the filtering process, the hand region is first extracted. Some portion of the forearm is presented in the same frame. Since the various lengths of forearm presented in the frame do not provide useful information to distinguish different hand postures. To do this, first estimate the width of the hand along the horizontal axis, and then record changes in width from right to left(from forearm area to hand) to locate the point with the largest width increase. Hand postures, only the hand region is used for classification. To extract the hand, first the wrist point is found. It works as follows: The frame is segmented into b_H bins along the horizontal axis as shown in fig-5, and the segmentation point which lies in between the *ith* bin and the *i+1* bin is denoted as ∂_i the size of the bin is $|max(x)-min(x)|/b_H$, where *max(x)* and *min(x)* represent the largest and the smallest x-address of events respectively, and b_H is the number of bins. Estimate the width of the hand for each bin, which is denoted by $d_{i,}$, for *ith* bin. Calculate the width change for each segmentation point ∂_i by subtracting the width of the hand in the left bin of ∂_i i.e d_i-$d_{i+1.}$ From forearm area to hand, find the segmentation point *p* with maximum changes. All the events whose x-address is smaller than *p* are regarded as a hand region.

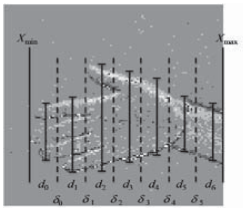


Figure5.Hand Extraction [1]

4.Feature Extraction

Feature extraction is very important in terms of giving input to a classifier. The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. One simple feature which can represent the shape of a hand might be the width distribution across the hand. Similar to the method to extract the wrist point, we segment the extracted hand into b_F bins along the horizontal axis as shown in the fig-6, and then estimate the hand width for each bin. A sequence of hand widths shows how the size of width changes along the horizontal axis. Since the absolute value of width can be different depending on person and the distance between the hand and the camera we use relative width, which can be obtained by dividing the absolute width of the bin by the sum of absolute widths over all bins. It is shown in equation 1, where relWidth(i) and absWidth(i) represent the relative and the absolute size of width at *i*th bin respectively.

$$\text{relWidth(i)=absWidth(i)}/\sum_{i}^{bF} absWidth(i) \qquad (1)$$

Similar to the process of hand extraction, the number of bins, b$_F$ should be appropriately set. If the number of bins is too small, key features cannot be captured. An extreme case is using one bin. Although computationally less expensive, the pattern of width and the hand shape cannot be inferred from only the average hand width. The opposite extreme case is using the number of column pixels of the abstracted hand data as b_F. Although the shape of hand can be represented at the fine-grained level, it is computationally expensive. More importantly, by focusing on too specific and too local patterns of data, we may fail to extract more general data patterns. If there are more possible gestures such as stretching fingers. Each finger can have only one of two states as stretched or folded, so that most of the commonly used hand gestures are combinations of the states of all five fingers. As a first step toward the recognition of the states of fingers, we attempt to recognize the number of fingers in each bin without distinguishing whether a finger is stretched or folded or by specifically identifying the fingers in each bin. To find the number of fingers within a bin, connected component analysis is used. Instead of filtering out the component with the smallest number of events as we did in the filtering process, the number of connected components for each bin is counted.

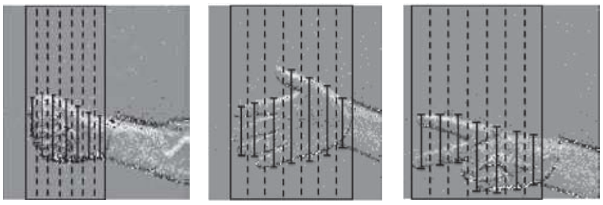


Figure6. The Frames at the Delivery Point

Next, as a step toward the recognition of the state of fingers, we attempt to recognize whether any of the fingers except the thumb is stretched or not. This could be easily recognized by comparing ratio of the length from the tip of fingers to the thumb to the length from the thumb to the wrist as shown in following fig-7. This can roughly estimate the location of thumb by locating the column with the maximum width. This feature is called as Horizontal Ratio.

IJECSE
International Journal of Electronics and Computer Science Engineering
WWW.IJECSE.ORG

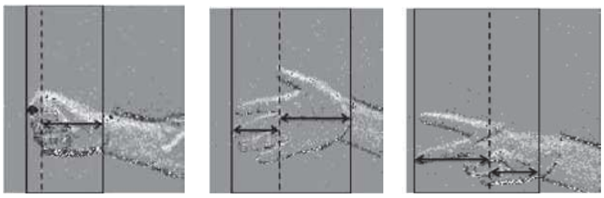


Figure7. Horizontal Ratios [1]

5.Classification

Once the features are extracted, a machine learning algorithm is applied to build a model for prediction. In this paper, we use naive Bayes algorithm which is a simple version of Bayesian network with an assumption that all the features are independent given the class. This method is an effective and fast method for static hand gesture recognition. This method is based on classifying the different gestures according to geometric-based invariants which are obtained from image data after segmentation; thus, unlike many other recognition methods, this method is not dependent on skin color. Gestures are extracted from each frame of the video, with a static background. The Naïve Bayes classifier works on a simple, but comparatively intuitive concept. An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters necessary for classification. In this width of palm and width of fingers are as input to Bayes technique and output will be classifying rock, paper and scissors hand gestures. It requires small amount of training data for classification. Training data for hand gestures are class features i.e. rock, paper, and scissors. Training data for mouse free interface are right click, left click and mouse click and class variables are rock, paper, and scissors as shown in following fig-8

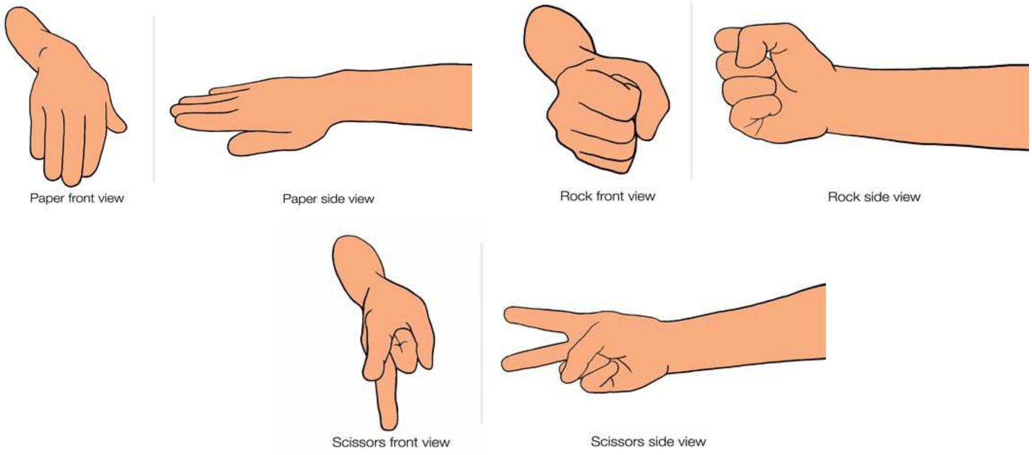


Figure8. Classification of Hand Gestures

III. EXPERIMENT AND RESULT

The proposed scheme is tested using ordinarily image processing. We used previously recorded DVS camera video data with two types of data. In one type of video date, a player delivers the same throw 20 times consecutively, for each rock, paper and scissors hand gesture. In another type of video, a placer delivers 40 random throws. In the random date, there are twelve rocks, fourteen papers and fourteen scissors, producing 40 rounds in total. In the previous section, we discussed two possible schemes of extracting features: width-based and component-based. Also, we introduce a feature called Horizontal Ratio which estimates the ratio of the length of finger to the length of palm. In this section, we compare the accuracy of classifiers when width-based and component-based methods are used for extracting features, and analyze the effect of using Horizontal Ratio on the classification accuracy. The first ten rounds of each rock, paper, and scissors date are used for training with manually provided labels. To test our model, we use the whole rounds of rock, paper, and scissors data and random date. Note that testing data with 100 rounds also includes training data with 30 rounds. Experimentally if we set TH = 7000, b_F= 15, b_F = 7, and the frame size = 20 ms Table 1 shows the accuracy of classifier for each feature extraction method in different date sets. The last column represents the average accuracy of the classifiers which is the ratio of total number rounds that gesture is correctly classified to a hundred rounds. Width-based Horizontal Ratio and component-based Horizontal Ratio methods use Horizontal Ratio as a feature in addition to seven features obtained by width-based or component-based method. In width-based component-based features extracted from both width-based and component-based are used with Horizontal Ratio, which produces 15 features in total. According to the results, the component-based feature extraction method slightly performed better than width-based methods, and Horizontal Ratio could contribute to the enhancement the classification accuracy. We expected that combining width-based and component-based method by simply concatenating the features from both methods can enhance the accuracy of classifier, but the results have shown it is not necessarily true. Although it performs similar or better in rock, paper, and scissors data its performance in random data was worse than that of component-based, component-based Horizontal Ratio or width-based Horizontal Ratio. The reason might be that the increase of the number of features leads the model to be over fitted to the training data.

Table -1 Classification Accuracies [1]

Methods	rock	paper	scissors	random	Avg
width-based	90	95	90	72.5	84.0
component-based	85	95	90	77.5	85.0
width-based + *Horizontal Ratio*	85	95	90.0	75.0	84.0
component-based + *Horizontal Ratio*	85	100	100	80.0	89.0
width-based +component-based + *Horizontal Ratio*	90	100	100	72.5	87.0

Our investigation on the process of feature extraction has shown that some frames at the delivery point contain so small number of events in the frame or around the boundary of objects, which makes it hard to recognize the shape of objects. It might happen if the player suddenly stops moving the hand or the size of hand appeared in the frame is so large compared to other data. We could also see that some portion of actual hand region is excluded as a result of hand extraction process by incorrectly detecting a wrist point. Such a problem is mainly caused by the player not putting his hand at the 90-degree angle at the upper body. However, our preliminary results were promising in a sense that about 89% level of accuracy was achieved by using Horizontal Ratio and component-based feature extraction method

IV.CONCLUSION

This paper proposes a method to classify bare hand gesture using dynamic vision sensor (DVS) camera. Specifically, we focused on classifying three different throws delivered by a player playing rock-paper-scissors game. We first described the properties of DVS camera which only responds to the pixels with luminance changes. Then, we proposed a method to detect a delivery point where the final throw is made. Once the delivery point is detected, only the still image of hand at the delivery point is used for classification. The region of hand is first extracted by locating the wrist point where the changes of the width of hand is the biggest, and then we extract the distribution of width within the hand or the number of connected components for each segment as features. The experimental results were promising; using component-based method and ratio of the length of finger to the length of palm results in 89% of the accuracy. Component-based feature extraction method performed slightly better than width-based method, and the ratio of the length of finger to the length of palm could contribute to the enhancement of classification accuracy.

However, there is a lot room for improvement. We discussed that the threshold for detecting delivery point should be adaptively changing depending on the size of actual object, and that the hand region should be more accurately detected in a situation where forearm is not at 90-degree angle at the upper body. It is expected that studies on detecting me finger state (i.e., open or folded) goes a step further. This work will be based on using the component-based method and it is expected to help make our system applicable in more general situations with numerous possible hind gestures. Furthermore, we plan to compare the classification accuracy and computational cost of using a DVS camera with those using conventional frame-based camera.

V. REFERENCE

[1] Eun Yeong Ann, Jun Haeng Lee, Tracy Mullen, John Yen,"Dynamic Vision Sensor Based Bare Hand Gesture Recognition",978-1-4244-9915-1/11/$26.00,2011IEEE

[2] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat," Vision Based Hand Gesture Recognition", World Academy of Science, Engineering and Technology 49 2009

[3] C.A.Bouman,"Connected Component Analysis", Digital Image Processing-9Jan,2012.

[4] Preeta Rajamani," Best Practices in Gestural Design", Bentley University

[5] http://www.worldrps.com/game-basics

[6] Snajay Meena,"A study on hand gesture recognition technique",Departement of Electronics and Communication engineering National Institute of Technology, 2011.