Head Gesture Recognition: A Literature Review

Er. Rushikesh T. Bankar
Ph. D Scholar,
Department of Electronics Engineering,
G. H. Raisoni College of Engineering,
Nagpur, India.

Dr. Suresh S. Salankar
Dean SAC & Professor,
Department of E&TC Engineering,
G. H. Raisoni College of Engineering,
Nagpur, India.

ABSTRACT

Head gesture recognition play the important role in research area. This paper shows a survey of head gesture recognition systems. Head gesture recognition system is mostly used for the peoples who cannot able to use their hands and legs mostly for the handicapped peoples. For head gesture recognition system the adaboost and the improved camshift algorithm i.e, the combination of camshift algorithm with BLBP (binary local block prediction) and UKF (unscented kalman filter) is used in this paper. There are various methods for head gesture recognition but here in this paper the eye template matching method is used. The PLK algorithm is used to detect the eye. Because of this eye template matching method exact position of the head is obtained. By comparing the initial and final window we obtained the position of the eye. After that the SVM (support vector machine classifier) is used. It is used for classification purpose, i.e, it classify the images into different groups (left, right, up and down).

Keywords - Adaboost Algorithm, Improved Camshift Algorithm (BLBP, UKF), Eye Template Matching Method for Head Gesture Recognition, SVM Classifier.

INTRODUCTION

In nowadays the head gesture recognition plays the important role in research area. This system is mostly used for the handicapped peoples, who are not able to use their hands and legs [1]. In today’s world Face tracking is mostly used method in research area ,it is used mostly for people who cannot use their hands or legs. The main application of this system is used in video surveillance, remote control for TV and etc. as there is a huge changes is occurred in environment conditions there is need to improve the tracking and detecting algorithm. The results can be improved by the combination of features [2]. Adaboost algorithm is used for head gesture recognition system. It is used as the face detecting algorithm. It is one of the most accurate algorithm used for the detection purpose. Extraction of features are done by the adaboost algorithm, in which the information of image is present only by integer calculation. The speed is high for detection purpose. And the camshift is used for tracking the face [3]. It gives the better result in color tracking. But it contains some disadvantages like (i) in illumination condition camshift algorithm gives the poor result; (ii) and also when the background is cluttered. The speed is fast but it fails in some environmental conditions. In this paper the solution is given to remove these disadvantages by doing the integration of camshift and adaboost algorithm [3]. Also there is need to improve the camshift algorithm by combining the camshift algorithm with BLBP (binary local block prediction) and UKF (Unscented Kalman Filter) algorithm. The result obtained from this improved camshift algorithm is better than that of the old camshift algorithm.it is able to track the faces in different environmental conditions. [11].

As there are different methods are used for head gesture recognition. But here in this paper the EYE template matching method is used, which denotes the position of the head (left, right, up and down). The PLK algorithm is used to denote the eye position. After that by comparing the initial window and the final window the eye position is obtained. From that we got the accurate head direction. It has good result in case of speed.
The multiclass SVM (Support Vector Machine) classifier is basically used for classification of the images. It is one of the machine learning algorithms. Analyzing the data is the main factor of SVM. Also, the detection of pattern is also the main factor. It is a supervised learning algorithm. It is divided into two parts as training and testing part. The training part of the SVM is used to generate the SVM and in testing part this generated SVM is used for classification of images.

The adaboost and camshift algorithm is describe in part II. The improved camshift algorithm is describe in part III. The integration of adaboost and camshift algorithm is given in part IV. The head gesture recognition using eye template matching method is given in part V. And the SVM classifier is given in part VI.

**ADABOOST AND IMPROVED CAMSHIFT ALGORITHM**

**A. Adaboost algorithm**

For head gesture recognition system adaboost algorithm is used in [3]. It is used as the face detecting algorithm. It is one of the most accurate algorithms for detection purpose. The detection rate of the adaboost algorithm is very high. It extracts some features and used for detection purpose. Against illumination conditions it is very strong. Adaboost algorithm detects the face under certain environmental conditions. But also it has some disadvantages. The block diagram is as shown below. It consists of the following steps:

- Data acquisition
- Pre-Processing
- Extraction of feature

![Block diagram of face detection algorithm i.e., adaboost algorithm](image)

this present time the human face detection method plays a very important role in real world and also to deal with the poor discriminability of haar like features. In paper [7] for a frontal face detection the PSO adaboost algorithm is used for tracking purpose. Which is based on the selection of multiple features. In this paper the PSO-Adaboost algorithm is proposed to increase the training process which is used to detect the face with the help of PSO (particle swarm optimization) technique. This paper proposed an efficient system of face detection which finds a frontal face with a cascade of classifier including generalized haar like descriptor. This PSO Adaboost is used to increase the generalization ability of feature selection and also increases the rate of selection of feature. It also decreases the training time.

A strong and popular face detection algorithms is needed for a system that gives the information of the face image. For that Yasaman Heydarzadeh [8] introduced a new face detection system, adaboost is used in the paper as the initial algorithm. Using Haar-like features for facial organs and gaining these by AdaBoost algorithm makes this system to find the occluded faces. Normally at initial level detection process contains some drawbacks: high false positive rate, which fails to finds the partially missing parts and poor detection rate. To solve this problems, several methods are applied as shown in Fig. 2.
the input image given is colored, skin mask is extracted, then the detection process is takes place only in extracted regions. Image enhancement is executed when input image is gray level, and after that the edge is find. And the Detection part is performed. If the value of the calculated edge for sub-image is in a specified range, then for detecting face this sub image will be send to next level. In order to restrict the search area these methods are used as initial stage and so speed up the approach. Face Detection and facial parts is given in the skin region. From this the facial parts are send towards the shown in the following fig. 3.

**Fig. 2. Diagram of face detecting processing [8]**

**Camshift Algorithm**

Camshift is one of the most popular algorithm to which is used to track the face. Here in paper [3] for obtaining the most accurate face detection, to track the object and real time gesture recognition the combination of camshift algorithm with the adaboost algorithm is introduced. In order to track the color of the face Camshift is a very useful method. The camshift algorithm is used for fast face detection According to this paper. But still it is not strong. It fails to track the face when the illumination condition occurs and when background is cluttered.

The result of both the Adaboost and Camshift algorithms are tested in this paper on different conditions: (a) Normal (b) Lighting (c) Darkness (d) noise in the Face color (e) Occlusion (f) Different color of face (g) Multi faces with different colors (h) more than one face having same colors. In contrast and noisy conditions, the result of Adaboost is very strong and faces are correctly detected. The result are shown in the below figure 4.

**Fig. 3. Detected face along with the eyes and nose [8]**

**Fig. 4. Result of Adaboost algorithm (row 1, 3) and the result of Camshift algorithm (row 2, 4) [3]**

Er. Rushikesh T. Bankar, Dr. Suresh S. Salankar
In order to overcome the limitations of camshift algorithm an improved camshift algorithm needs to develop. For real-time accurate target tracking, Saket Joshi [9] introduced two algorithms as Camshift and SURF algorithms. For some environmental factors Camshift is very sensitive and it uses the color features. Also it takes help from information which obtained from color to track the target when it is moving. For showing the color history in the image, the color histogram is used. When exactly same colors are occurred in the background, old Camshift algorithm is not properly tracked the object. Also Camshift is not strong in the case of background noises.

To decrease the illumination limitations a new Camshift algorithm is introduced. In order to fit in different illumination cases, the new Camshift is able to automatically change its factors but still some drawbacks are there to overcome. It is seen that the system is non-ideal, when background contains the same objects. Second limitation is that an object is not very hard to lose under dynamic background because Camshift uses Hue feature. The accuracy is improved when it is added to Camshift. As SURF has much higher computational complexity only limited use is possible only to refining the object [9].

To solve a limitations of old CAMSHIFT in paper [10] proposed the combination of the camshaft, kalman and adaboost algorithm. CAMSHIFT is a semiautomatic tracking method and also it is a color-based method. But again, two limitations is required to be improved: 1. the poor results are obtained for moving object. 2. When the same color is present in background it fails. For implementing this method the open source code library is used. In the detection part, Haar features are used to train the face classifier which already stay in an open source-code library. By using the Haar-based face classifier Haar features are useful in recognition system. By using the Haar-based face classifier, the speed of the detection is very fast. The position and size of face is send to the CAMSHIFT, which will be analyzed and convert it into RGB space to HSV space. Then the color probability distribution is calculated, projecting to the back-project image. According to this CAMSHIFT tracks the object. Before they start tracking, a morphology filter is used to remove the noise and increase the accuracy of tracking. This method find the solution on track loss drawback also for moving target motion and noise. It is very robust against noisy conditions.

**IMPROVED CAMSHIFT ALGORITHM USING BLBP AND UKF ALGORITHM**

The behavior and the result of the camshift is not so good. It fails in various situations. Also when there is a change in outside conditions. A new camshift algorithm corrects the result of the camshift algorithm. The new one is nothing but the BLBP and UKF algorithms.

**Improved Camshift algorithm using BLBP Algorithm**

In paper [11] the new algorithm is designed with the BLBP algorithm, for correcting the limitations. This algorithm is not used for many times. So this paper gives the review on this algorithm. The latest method is none other than the camshift with BLBP. The good results are obtained.

If the same things is present in back side, then this is very easy for this algorithm to remove such things from background. To increases the quality of tracking, this algorithm is used. It is not so complicated method. Complexity removal is very easy task for this algorithm.

**Improved Camshift algorithm using UKF Algorithm**

Zouguo Yan [12] uses the UKF with the camshift, for getting the best future tracking. Facial characteristics plays important for this. They also used the kalman filter. Also in some cases surf is used as when more than two object are present.

The block diagram of the camshift and ukf is as shown below in fig. 5. In that first the videos or images need to be provided to the new algorithm. The image is converted into the RGB to HSV. After that it observed the UKF algorithm. Also it is beneficial in case of prediction. After that the final image is tracked. This algorithm is used for quality improvement. The proper result is as shown below.
The algorithms, like camshift, meanshift and kalman are used in paper [13]. This three algorithms find the object and after that tracking is done. These three algorithms gives the best result in terms of accuracy and decrease the consumption of time.

Xianggong Hong [14] gives the new algorithm over the old one. From which the good results are obtained. Lbp histogram is also obtained. The binary code of the image is obtained by comparing with its pixel. The whole process is required for improving the result. At last the kalman is also used to improve the results. The result is as given below:

Gi-Woo Kim [18] tell us something about the old and the new algorithm. The old one is the camshift and the new one is the improved camshift i.e. the the new one is with the kalman filter. It gives the good result.

The LQE contains the same factors as that of the kalman filter. The noise is stored in that algorithm. So that by skipping the noise, the new algorithm is found for tracking. Which is shown in the following block diagram. First input image is gave after that some preprocessing takes place. After that the kalman and camshift got the object. And tracking is done.

Fig. 5. Block diagram of camshift algorithm combined with UKF algorithm [12]

Fig. 6. The result of improved camshift algorithm with UKF algorithm [12]

Fig. 8(a). With kalman [14]

Fig. 8(b). Without kalman [14]
INTEGRATION OF ADABOOST AND CAMSHIFT ALGORITHM

The face is found by using the adaboost algorithm [3] and then as usual camshift is used. The satisfied results they got and the whole process is done very fast. But it is very weak for getting the result. Integration is the only way for this limitations. The block diagram of integration of adaboost and camshift algorithm is as shown below [3]

HEAD GESTURE RECOGNITION USING EYE TEMPLATE MATCHING

As we all know the eyes are the good example in case of gesture recognition. So for that some information is obtained from paper [4] for gesture recognition. The eye template matching is used for this. Which gives the good result. For that the detection task is carried out first.

And the main task after the detection is the tracking. For tracking, in this paper PLK algorithm is used. It gives the good result.

The human being has two eye pupils. Tracking this eye pupil is the main task. The calculation process shows as follows in Fig. 10.
From the above figure they got the initial frame. And after doing some process they got the final frame. So by using this two frames they got the position of the eye. If it shows the poor result then OTSU algorithm is used to correct the result. It is used to obtain the threshold of the binaries image. It is one of the best method for head gesture recognition.

Nilamani Bhoi [15] also gives us the knowledge about the same eye template matching method. It is very beneficial in case of security, laziness and etc. it gives the good result.

The diagram for the further process is as shown above in fig. 11. As we all know that the overlapping is the main factor for

Finding the features. For that first selection of face is takes place after that they calculate the 2D correlation. After that they find the mean square error. From which they got the eye region. And obtained the accurate result i.e. position of the eye. Not only it gives the proper result for open eyes but also gives the result when the people is in sleeping mode i.e. for closed eyes. The results are as shown below in the fig. 12.
The paper [16] introduced the method depends on image which is consists of three types which is as follows the feature based method, template based method and the last one is the appearance based method. The characteristics of eye is found out in first method for finding the eye features. In template matching method the part of the eye is developed which is look like the original eye part. After that it tries to match that part with the original part from that they got the position of the eye. But it is not the time saving method. Also we can find the part of the eyes by using the characteristics features. Then whatever data is obtained is used for classifier for classification.

The template matching which depends on LEM is applied in the first method for finding the eye region. The two midpoints are obtained by using the last method which are iris midpoint and the eye midpoint. The results were improved by using the methods which are used in this paper.

M.J.T. Reinders [17] presents tracking algorithm for an eye which is strong for changes in scale, orientation and eye gesture, like blinking of eye. The template matching method is used to locate the eye regions in the different frames. This technique is strong for changes in the eye parts. Here, this paper proposed algorithm basically for tracking which is based on template matching. After the initialization new eye templates learns automatically, placed in a codebook which is strong enough for variations in eye parts. The template matching has an iconic view (the ‘template’) that contains maximum matching to that of other images. Some parameters has a vital role in template matching method which are the iconic view or the set of iconic views, and the similarity measure.

According to the paper when template matching method is used, for tracking the location of part of eyes to an image sequence is very straightforward. Similarity of the positions measure peaks within a single frame, then determines the positions of the sought characteristics. When the eyes may be blinking, the size of the eyes changes. As template matching is not strong for shape variations, only single representative iconic image is used. It is very well conceivable. From this mismatches and consequently error accumulation occurs. In this paper to create the initial template they select the manual intervention, i.e. the user should notice the midpoint of the left and right eye parts. The initial templates for the left and right eye which is based on the physiognomy are a created by its own. The result of head movements, should be matched for changes in orientation and the eye parts. This paper proposed an eye location estimation method. This consumes the more time for that this paper make use of search windows (regions of interest). It increases the strongness of the eye tracker. It explain a scheme for tracking the left and right eye regions throughout an image sequences. But, as per given in the paper as, only one iconic view for each eye region is available. This cause the various problems when the eyes are tracked, also the illumination is not constant over the time lapse of an image sequence. For both the changes Template matching is very sensitive. For these limitations this paper proposed to use a codebook of iconic views of that region. During the feature region tracking process the distinct iconic views are learned automatically. This tracking scheme is little bit accurate except variations in orientation, scale and shape of the feature region.

**SUPPORT VECTOR MACHINE CLASSIFIER**

Support vector machine is nothing but the artificial learning method. It is used in case of data analyzing. It is used in classification purpose. It is consists of the two types the first one is the training part, where the svm is generated. And the second part is the testing part, in which the generated svm is used to classify the images.

The preprocessing and the extraction feature based hand gesture recognition is given in [15]. The preprocessing consists of the image enhancement and segmentation. As per the name the feature extraction extracts the feature. Then the classifier is used for classification purpose by using these features. It gives the best result when the features are combined.

M.Murugeswari [19] implements the hand gesture recognition which depends on the vision. Which controls the robot movement? This process is done by using the SIFT algorithm. It extracts the feature. The main working flow chart of this paper is as follows:
In the above diagram first the camera is used to capture the image. It captures the gesture. After that it converts into PGM format. Then by using SIFT algorithm extraction of feature is done. After that it generates the bag of world. And the obtained features are then used for classification purpose. Classification is done by the training and the testing method. At last the command is generated which is used to control the robot movement.

Zhu Weixing [20] used the classifier to classify the different gesture of procine body. The gesture may be of different types, according to the paper the classification is performed using the three different gesture like when it stands in normal condition, standing with drooped head and lying, based on the training and the testing method. It is also used to find the mental state. The procine body with different size is also shown in the following figure. 14.

![Image Sets](image-url)

**Fig. 14. Image sets [20]**

It decides the position of the swine depends on its ellipse center. When it is in second quadrant then it is in standing with head in up position. In third quadrant it shows the position with head droopy. And in the fourth quadrant, it shows lying position. The result is as shown below:

![Fig. 15(a).](image-url)

**Fig. 15(a). Standing with normal spirit [20]**

![Fig. 15(b).](image-url)

**Fig. 15(b). Standing with spirit droopy [20]**
The system which is used in paper [21] is dividing into three states:
- The dataset making state
- Training set and
- Prediction state

The diagram is as shown below:

Fig. 15(c). Lying [20]

Fig. 16(a). Data set making steps [21]

Fig. 16(b). Train steps [21]

Fig. 16(c). Prediction steps [21]

Fig 16(a) denotes the first state, the hand images is captured in various gestures. After that it selects the region of interest of the hand. Then we have to resize the image by 40*40 pixel. After that for color factor the image is converted from YCBCR to BGR color space. Filtering is needed in case of any image contains the noise. This process reduces the noise. After that the final result is obtained. The obtained result is noise free and then it stored the result and again it jumps on the first step.
The main working of the second step i.e. training step is used to find the HOG descriptor. It divides the data into training and testing state which is as shown in the fig. 16(b).

And in the last state, which is shown in fig. 16(c), the frame is captured after that the image is converted into YCBCR. From this they found the face. Then this image have to be resized into proper scale. After that this is converted into grayscale for obtaining the HOG descriptor. Descriptor also needs to be resized. And finally the result is obtained.

In this paper the SVM is used to recognize the pattern in terms of image manipulation. For this process different videos and images of hands with different gesture is taken. After that there is a need of histogram for feature extraction. Later it is given to the classifier for classification [21].

II. CONCLUSION

In this paper head gesture recognition is describe by using the different methods. In this paper the study of adaboost algorithm is given. It is mainly used for face detection. The result obtained from this algorithm shows the accuracy [3]. For tracking the camshift is used [8]. As it contains the advantages it also has some disadvantages like the results are failed in illumination condition and if the background is cluttered. For getting the better result integration of the camshift and the adaboost is the only way [3]. To improve the results of old camshift algorithm an improved camshift algorithm is introduced i.e. the combination of camshift with BLBP and UKF. The BLBP (block local binary pattern) gives the better result than the old algorithm. It combines with the color features. The result is more robust. [11]. UKF (unscented kalman filter) is used to denote the position. For that this algorithm is combined with the camshift algorithm [12]. Then this paper focus on the eye template matching method for head gesture recognition system. For tracking, the PLK algorithm is used [4]. Location of the eyes are found by using the template matching [17]. And after that the classifier is used i.e. the SVM. It is of two parts the first one is the training part which is used to generate the SVM and the second one is the testing part which is used for classification [15]. The result obtained by using this method shows the accuracy [21].

REFERENCES


101

Er. Rushikesh T. Bankar, Dr. Suresh S. Salankar