Hand Gesture Recognition using multi-objective optimization-based segmentation technique

*[1]Supriya.S, K.S.R College of Engineering,Tiruchengodu,mail:ssupriyaa85@gmail.com
[2] Dr.Manoharan.C, Sapthagiri College Of Engineering,Dharmapuri,TamilNadu, mail:c_m_66@yahoo.co.in
*Corresponding author: ssupriyaa85@gmail.com

Abstract:

Hand Gesture Recognition (HGR) software is winding up progressively open with the advances in depth cameras and sensors, however, these sensors are as yet costly and not uninhibitedly accessible. A continuous HGR programming is intended to work with a minimal effort monocular web camera. Skin discovery and skin extraction is a typical type of image handling utilized for motion acknowledgment. The hand gesture image is gone through three phases, preprocessing, feature extraction, and characterization or segmentation. In the preprocessing stage, a few tasks are connected to separate the hand gesture from its experience and set up the hand gesture image for the feature extraction stage. In this paper, Multi-Target Optimization Based Segmentation (MTOBS) has been proposed for HGR. The performance has been analyzed for gesture recognition with and without optimization technique. The outcomes demonstrate that the recognition method without optimization has an exhibition of 85% recognition, while the proposed technique with optimization, has a superior execution of 96% recognition rate.

Keywords: Image Processing, Image Segmentation, Feature Extraction, Gesture and Zernike Moment.

1. Introduction

Gesture recognition is a type of perceptual computing user edge that allows computers to capture and take human gestures as commands. Gesture identification based on computer vision technology has been received great interests recently, due to its natural human-computer interaction characteristics. The advantage is, user can control the devices without touching the keyboard, panel or mouse. Hand gesture method gives more freedom for the user to control the device like television remote control, presentation slide control is some of the applications of this hand gesture. The main drawback is hand motion speed is ever fast and complicated compared to the computer image processing speed and people may need more training for this version. The general framework is first finding the location of the input image and it's segmented which is depends on the skin color of the image. The second step is finding the motion of the hand.

There are a lot of difficulties in accurate hand gesture recognition. The obstruction could increase the difficulty in pose recognition. The use of hand gestures provides an attractive and natural substitute to these bulky interface devices for human-computer interaction. Using hand as a device can help people to communicate with computers in an inherent way. The most effective tools for capturing hand gesture are electromechanical or magnetic sensing devices. In this, the sensors are enclosed with the hand gloves that transduce finger flexions into electrical signals to define the hand gesture but it has a drawback it hiding the naturalness of the hand. Hand gesture recognition systems detect and segment hands using marker-aided methods. But these methods are difficult when compared with markerless vision-based solutions.

The organization of the paper is as follows; the detailed explanation of hand gesture recognition is given in section.1. Section 2 describes the various linguistic description of the hand. Section 3 presents the proposed method for gesture recognition. Section 4 gives the experimental results. Section 5 described the findings and discussions and finally, the conclusion is given in section.6.

2. Linguistic description of the hand

The regular linguistic description gives data about significant patterns and a normal practice present in datasets, however, they additionally protect data about their representativity, by methods for verb modifiers demonstrating the degree to which the patterns are normal for the entire datasets. The hand arrangement will be characterized by nine phonetic factors. Most commonly, two factors are utilized for the thumb portrayal, they are,

Thumb Configuration-It portrays the extending of the thumb.

Thumb Orientation-It portrays the direction in the hand reference outline.

Four factors are utilized for the portrayal of different fingers called the long fingers: the *finger Configuration*(i=2,3,4,5). They give the shape of the relating long finger and are also named *index, middle, ring, and pinkie.* The last three variables are utilized for the portrayal of the overall dispersing of fingers 2/3,4/3 and 5/4. They are individually named *finger Abduct$_i$ (i = 2,...,4)* or*list Abduct, ring Abduct, and pinkie Abduct.*. The French communication via gestures Alphabet and Numbers indication of Hand gesture is given in figure.1.

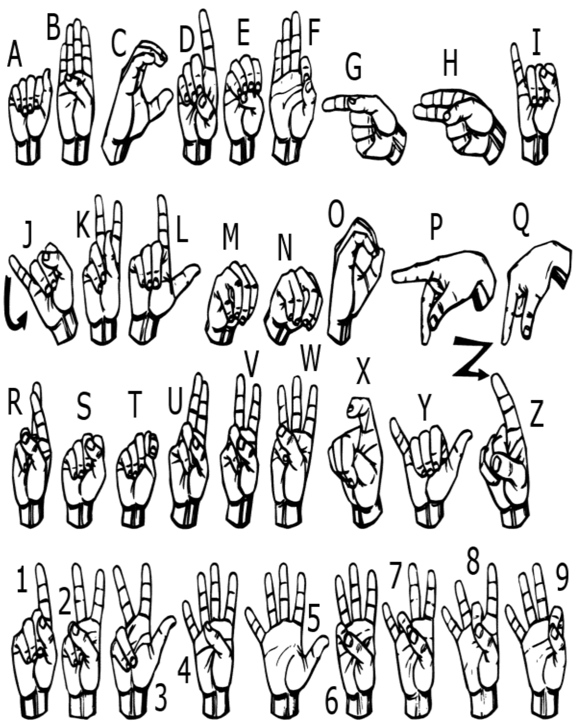


Fig.1 French communication via gestures Alphabet and Numbers

3. Proposed method

The hand gesture recognition using MTOBS is given in fig.2. Here, the gesture images from the standard dataset “BochumGestures1998” has been used as the input for the proposed system. Initially, the gesture images are given to the pre-processing block. Where the required images are subtracted from background and the noise present in the gesture images are filtered by using a median filter. Then, the feature such as Zernike moment has been extracted and optimized using Multi-Target optimization technique. These optimized features are used to recognize the gesture from the hand images.

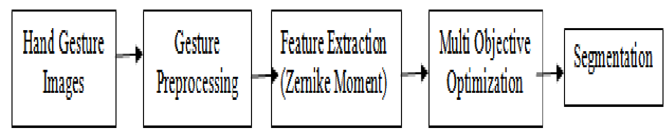


Fig.2. Hand Gesture Recognition using MTOBS method

3.1 Hand Gesture Dataset

In this work, publicly available dataset “BochumGestures1998” has been used. This dataset has static hand gestures, collected at the Institute fuer Neuroin formatik, Ruhr-Universitaet Bochum, Germany. All images from this dataset are 128x128 tiff images. The hand posture is identified by a code "00" to "12" in the filename. The kind of background light, dark, complex) is indicated by a subsequent letter a, b, or c, respectively.

3.2 Gesture Preprocessing

The main objective of this block is to locate and isolate the feature vector points from the hand gesture patterns. Sachin and Apeksha [14] have demonstrated a moving cursor in PowerPoint slide with the help of hand gestures. Consequently, the Zernike moment is used to determine the feature points of an object within a unit circle with respect to rotation and shift invariant points. It is claimed that the system consists of 100 frames with a difference in its orientation and position.

3.3 Zernike Moment

The displaying of hand portrayal in various measurements depends on the organizing frameworks. In existing procedures, the element can't be perceived appropriately because of its shape depiction, for example, linearity, circularity, and so forth inside a picture. Consequently, the framework utilizes Zernike minute to get higher request highlight focuses in space arrange of request N with various minutes. Give us a chance to consider a mind-boggling set-in Zernike minute polynomial which satisfies the symmetrical property inside the roundabout limit of an article. It is defined as $a^2+b^2=1$, where a and b are rectangular organize tomahawks. Zernike minute an of 'm' request and 'n' dreary stage is given in Eq. (1).

$$A_{mn} = \frac{m+1}{\Pi} \oiint_{a^2+b^2 \leq 1} Z_{mn}(\gamma, \varphi) g(a, b)\, da\, db \quad (1)$$

Where *g (a, b)* represents intensity of the image at *(a, b)*, *m* represents positive integer, n is integer, if $m-|n|$ is even integer and $|n| = m$, $a^2+b^2=1$. ϕ is he angle between a-axis in anti-clockwise direction. $Z_{mn}(\gamma,\phi)$ is a complex conjugate of $B_{mn}(\gamma)$. Radial function is defined as sfollows,

$$\boldsymbol{B_{mn}(\gamma)} = \sum_{c=0}^{m-n/2} \frac{-\boldsymbol{1}^{b}[(\boldsymbol{m}-\boldsymbol{b})!\,\boldsymbol{\gamma}^{m-2b}}{\boldsymbol{b}!\,(\boldsymbol{m}+\frac{|n|}{2}-\boldsymbol{b})(\boldsymbol{m}-\frac{|n|}{2}-\boldsymbol{b})} \quad (2)$$

When an article is situated concerning focus of circle, some pixel esteem drops out the unit circle which can't be represented figuring. As *Bmn* is the perplexing variable of motion designs. Zernike minute satisfies rotational invariant as a scalar amount. It gives preferable element focuses over other shape descriptor approach. Before extricating highlight focuses, it is imperative to accomplish scaling and revolution invariance by standardization and picture interpretation as recommended by Gholam Reza et al. [15].

3.4 Multi target Optimization Technique

Multi-target optimization considers optimization problems involving more than one target function to be optimized simultaneously. A solution is called non-dominated or Pareto optimal if none of the objective functions can be improved in value without degrading one or more of the other target values. In multi-objective optimization problem, the goodness of a solution is determined by the dominance. The multi target optimization technique is shown in fig.3. The proposed optimization algorithm has the following steps,

Step-1 Get the hand gesture images from the BochumGesture1998 dataset and the dataset is randomly divided into RGB and SIH subsets.

Step-2 Get the extracted features form a dataset.

Step-3 Initialize the genetic population P 0 of size N and to initialize the threshold value t=0.

Step-4 If *t* is less than maximum iterations, then the threshold value is increased otherwise the Population Q_t of size N is obtained from P_t through crossover and mutation.

3.5 Segmentation

Region Growing (RG) based segmentation method has been used to segment the hand gesture images. This method presents some hazy impacts amid segmentation process, so it prompts over segmentation. To conquer this issue, GWO technique has been utilized to optimize the threshold value which is acquired from RG method.

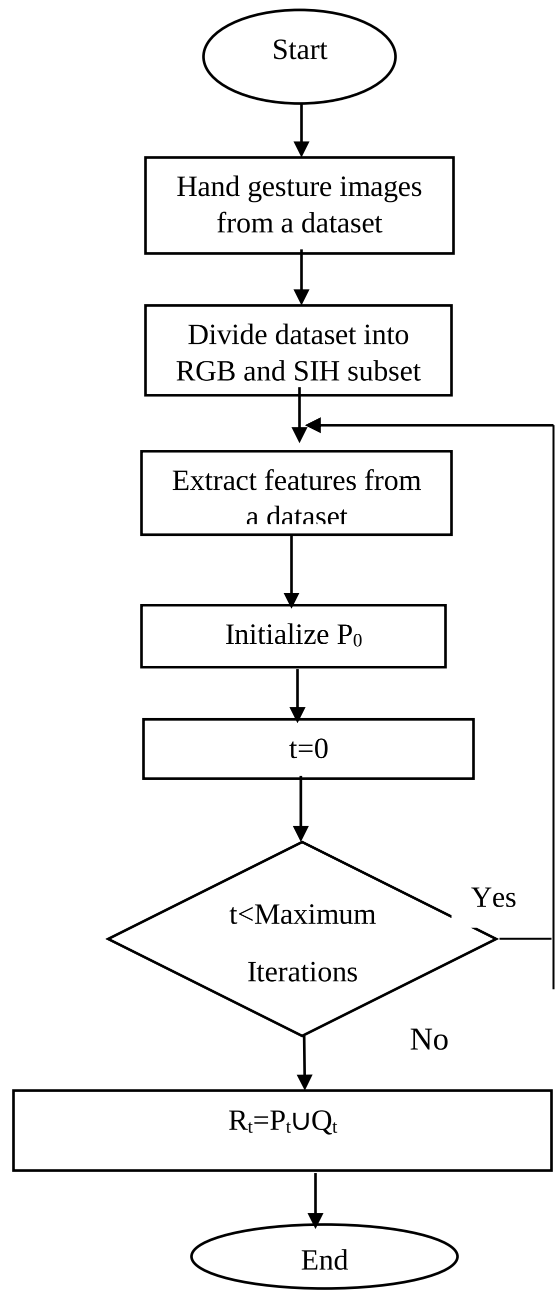


Fig.3 Multi-Target Optimization Technique

3.5.1 Gridding

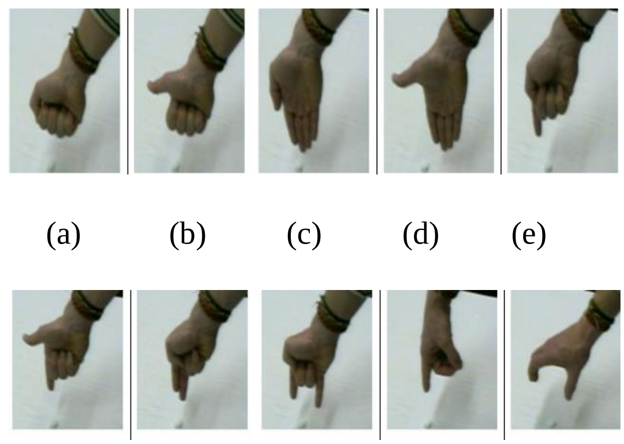
Gridding is realized to discover a progression of even and vertical lines to isolate one slide image into sub-matrices and individual spots territories. The system of segmentation isolates the spot pixels into closer view, foundation, or commotion. At long last, the intensity extraction is gone for getting the quality articulation levels as per the past operational outcomes.

3.5.2 Seed point selection (Region Centre)

The pixel values of the original image lie in the range of 0 to 255. The most frequently happening pixel value is assigned to be the seed point. Histogram equalization method is used to find the seed point of the image. Then the threshold value of the hand gesture images is calculated. This value is used to segment the hand from the hand gesture images. In order to improve the segmentation accuracy, the obtained threshold value must be optimized. This threshold optimization is performed by using GWO optimization technique. Then, the obtained optimized value is used for segmentation process. So that, we can get the segmented image.

3. Experimental Results

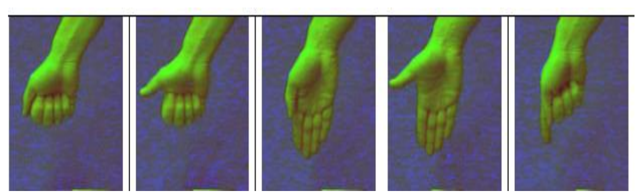
In this paper, only 20 images of the dataset have been shown. Among this, 10 RGB images and the remaining 10 are belongs to SIH images. The images which are used for this experiment is shown in figure.4 and 5. Fig 4 Shows the hand gesture images in RGB domain, and fig.5 shows the input hand gesture images in SIH domain.



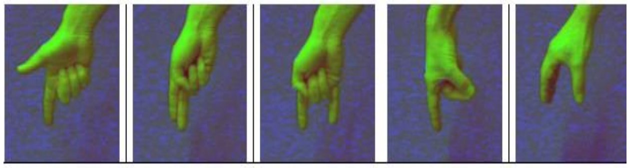
(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 4 Input Hand Gesture Images in RGB domain



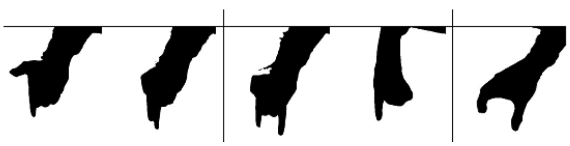
(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)

Fig. 5 Input Hand Gesture Images in SIH domain

(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)

Fig.6 Segmented Hand gesture images (RGB) without optimization

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig.7 Segmented Hand gesture images (SIH) without optimization

The segmented results using proposed method without optimization is shown in fig. 6 and fig.7. Fig.6 shows the segmented hand gesture images in RGB domain and the segmented results in SIH domain is given in fig.7.

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig.8 Segmented Hand gesture images (RGB) with optimization

(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig.9 Segmented Hand gesture images (SIH) with optimization

The segmented results using proposed method with multi objective optimization is shown in fig. 8 and fig.9. Fig.8 shows the segmented hand gesture images in RGB domain and the segmented results in SIH domain is given in fig.9. By observing the segmented hand gesture images, it is clearly noticed that, the segmented results multi objective optimization technique is better than the segmented results without optimization technique.

5. Performance Evaluation

Due to the dimensional variations of each database images has been independently evaluated to find the accuracy of the proposed method. The output of each dataset has been initially converted to binary images and compared with the ground truth images. Here, the most effectively used evaluation parameters such as accuracy, selectivity and sensitivity has been used for performance evaluation. These values depend on the TP, FN, TN and FP values. The below equations (2-4) has been used to find the parameters, accuracy is given by [16]

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} x100 \quad (3)$$

Another performance metric specificity is calculated by using the following equation [17], the value of specificity must be high for better segmentation.

$$Specificity = \frac{TN}{TN + FP} x100 \quad (4)$$

The next important parameter is sensitivity, it is expressed as below equation [18]. The segmentation method produces better result when the sensitivity value is high enough.

$$Sensitivity = \frac{TP}{TP + FN} x100 \quad (5)$$

Where, TP represents True Positive, FN represents False Negative, TN represents True Negative and FP represents False positive respectively The evaluation parameter values for segmentation accuracy, sensitivity and specificity are listed in table.1 and table.2.

Table.1 Evaluation of Segmentation Accuracy, sensitivity and specificity with and without optimization for hand gesture images in RGB domain

Images	Segmentation Accuracy		Sensitivity		Specificity	
	With Opt	Without Opt	With Opt	Without Opt	With Opt	Without Opt
01	0.98	0.92	0.98	0.85	0.96	0.84
02	0.99	0.84	0.98	0.83	0.96	0.87
03	0.9	0.85	0.9	0.84	0.9	0.84

	8		6		5	
04	0.96	0.85	0.95	0.82	0.95	0.83
05	0.94	0.84	0.98	0.82	0.93	0.82
06	0.92	0.83	0.92	0.83	0.92	0.88
07	0.96	0.91	0.94	0.84	0.98	0.87
08	0.94	0.84	0.96	0.85	0.92	0.85
09	0.93	0.81	0.95	0.87	0.96	0.82
10	0.94	0.86	0.95	0.86	0.95	0.83

Table.2 Evaluation of Segmentation Accuracy, sensitivity and specificity with and without optimization for hand gesture images in SIH domain

Images	Segmentation Accuracy		Sensitivity		Specificity	
	With Opt	Without Opt	With Opt	Without Opt	With Opt	Without Opt
01	0.96	0.85	0.98	0.85	0.97	0.84
02	0.96	0.88	0.95	0.83	0.95	0.85
03	0.95	0.86	0.96	0.82	0.96	0.84
04	0.95	0.85	0.94	0.84	0.94	0.87
05	0.92	0.84	0.92	0.85	0.95	0.83
06	0.94	0.82	0.97	0.85	0.98	0.81
07	0.92	0.82	0.98	0.84	0.97	0.82
08	0.94	0.86	0.92	0.83	0.95	0.84
09	0.93	0.87	0.93	0.88	0.92	0.87
10	0.97	0.85	0.91	0.87	0.95	0.88

Table.1 gives the performance values for proposed method in RGB domain, and table.2 gives the performance values of proposed method in SIH domain. This performance evaluation table clearly shows that, the segmentation accuracy, sensitivity and specificity is far better in MTOBS when compared without optimization technique. The proposed MTOBS method produces better segmentation results in terms of average segmentation accuracy of 96%, sensitivity of 95% and specificity of 95%. The segmentation accuracy of proposed method without optimization is 85%. Therefore, the accuracy of the proposed method is increased by 10%.

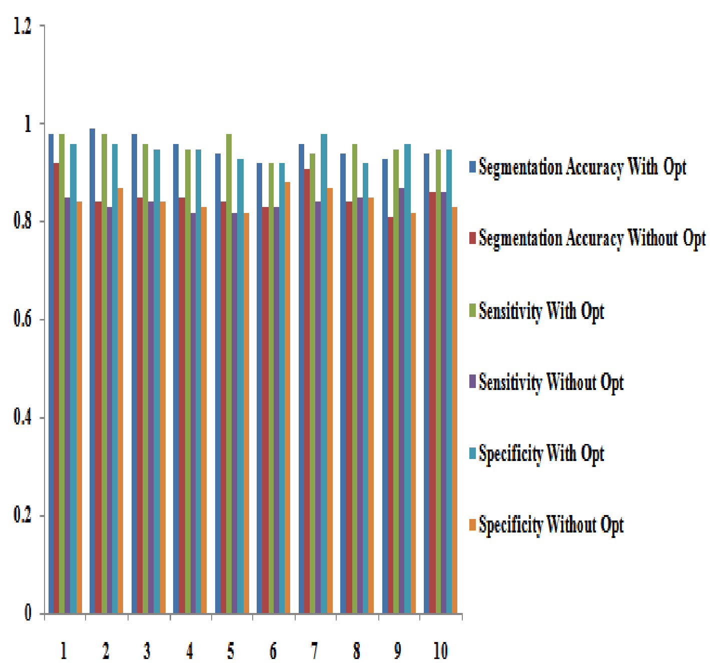


Fig. 10 Comparative Analysis of performance parameters of segmented

Hand gesture images (RGB) with and without optimization

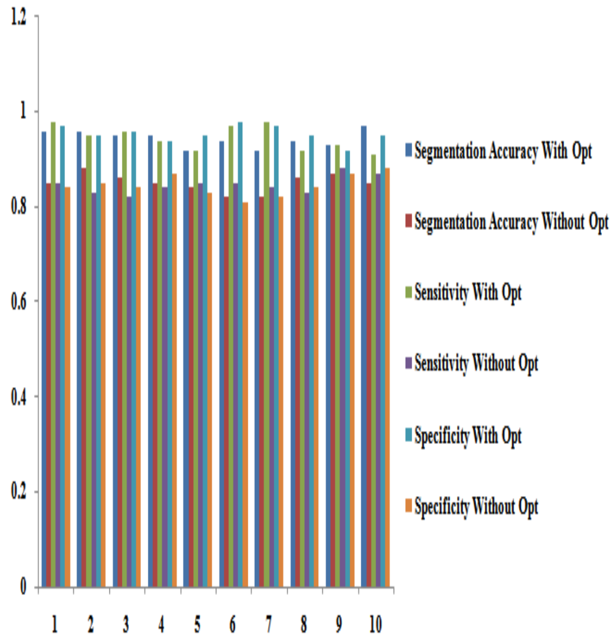


Fig. 11 Comparative Analysis of performance parameters of segmented Hand gesture images (SIH) with and without optimization

The comparative analysis of performance parameters of segmented Hand gesture images (SIH and RGB) with and without optimization is given in fig.10 and fig.11. From the analysis, it is clearly noticed that, the proposed method produces better segmentation results with the segmentation accuracy of 96%. Therefore, it is strongly concluded that, the proposed method is well suitable for segmenting the hand gesture images for hand gesture recognition.

Table.3 Recognition Accuracy of real-time gesture data for various methods.

Methods	Year	Dataset	RA (%)
H. Ragheb et al., [1]	2008	ViHASi	72.00
C. C. Chen & J. Aggarwal, [2]	2009	UT-Tower	90.43
M. S. Ryoo & J.K. Aggarwal [3]	2009	UT-interaction	91.67
J.C. Niebles et al., [4]	2010	Olympic sports	91.10
A. Patron-Perez et al., [5]	2010	TV human interaction	46.00
H. Kuehne et al., [6]	2011	HMDB51	57.20
G. Denina et al., [7]	2011	Video we	72.00
K. Soomro et al., [8]	2012	UCF-101	83.50
K.K. Reddy & M. Shah, [9]	2013	UCF-50	91.20
M.M. Gharasuie, H. Seyedarabi [10]	2014	Numbers (0-9)	93.84
H. Kim & I. Kim, [11]	2015	Kyonggi dataset	95.00
Archana Ghotkar et al., [12]	2016	ISL (20 dynamic signs)	89.25
Pablo Barros et al.,[13]	2017	RPPDI	93.33

Table.3 gives the recognition accuracy of real-time gesture data for various methods. This values has been used to compare the efficiency of the proposed method. The proposed method gives the recognition accuracy of 96%. This comparative analysis is depicted in fig.12

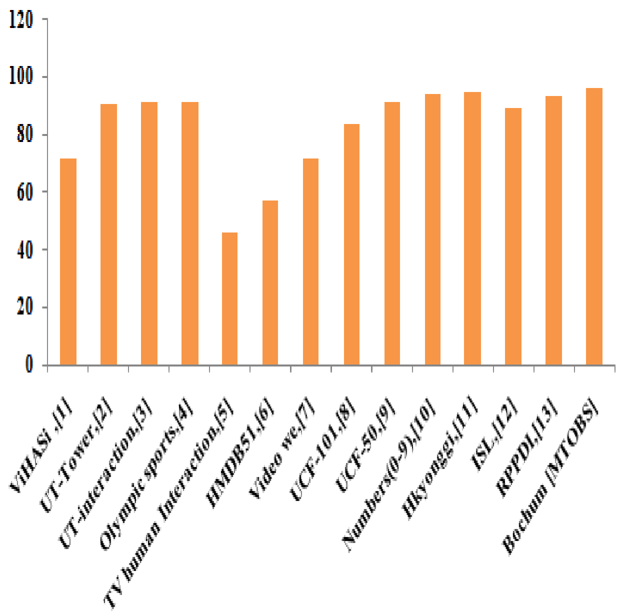


Fig.12 Comparative Analysis of Recognition Accuracy with existing other methods

6. Conclusion

In this paper, an optimization-based segmentation approach is proposed for hand gesture recognition. Multi objective algorithm is used for optimization. The work is carried out for BochumGestures1998 dataset. Experiments are carried out on segmentation with and without optimization. The average value of recognition rate with multi objective optimization has been increased by 10%. The proposed system has also yielded higher recognition rate than the other conventional hybrid techniques used in hand gesture recognition system.

References

[1] H. Ragheb, S. Velastin, P. Remagnino, T. Ellis, Vihasi, Virtual human action silhouettedata for the performance evaluation of silhouette-based action recognitionmethods, 2nd ACM/IEEE International Conference on Distributed Samrt Cameras (ICDSC) (2008) 1–10.

[2] C.-C. Chen, J. Aggarwal, Recognizing human action from a far field of view, Workshop on Motion and Video Computing (2009) 1–7.

[3] M.S. Ryoo, J.K. Aggarwal, Spatio-temporal relationship match: video structurecomparison for recognition of complex human activities, IEEE InternationalConference on Computer Vision (ICCV) (2009) 1593–1600.

[4] J.C. Niebles, C.-W. Chen, L. Fei-Fei, Modeling temporal structure of decomposablemotion segments for activity classification, European Confernce on ComputerVision (ECCV), Springer, 2010, pp. 392–405.

[5] A. Patron-Perez, M. Marszalek, A. Zisserman, I. Reid, High five: recognizing humaninteractions in TV shows, Proceedings of the British Machine Vision

Conference (BMVC) (2010) pp. 50. 1–50.11.

[6] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, Hmdb: a large video databasefor human motion recognition, IEEE International Conference on Computer Vision (ICCV) (2011) 2556–2563.

[7] G. Denina, B. Bhanu, H.T. Nguyen, C. Ding, A. Kamal, C. Ravishankar, A. RoyChowdhury, A. Ivers, B. Varda, Video web dataset for multi-camera activities andnon-verbal communication, Distributed Video Sensor Networks, Springer, 2011, pp.335–347.

[8] K. Soomro, A.R. Zamir, M. Shah, UCF101: A Dataset of 101 Human Actions Classesfor Videos in the Wild, Technical Report CRCV-TR-12-01, University of CentralFlorida, 2012 November.

[9] K.K. Reddy, M. Shah, Recognizing 50 human action categories of web videos, Mach.Vis. Appl. 24 (5) (2013) 971–981.

[10] M.M. Gharasuie, H. Seyedarabi, Real-time dynamic hand gesture recognition usinghidden Markov models, 8th Iranian Conference on Machine Vision and ImageProcessing (MVIP) (2014), http://dx.doi.org/10.1109/IranianMVIP.2013.6779977.

[11] Hyesuk Kim, Incheol Kim, Dynamic arm gesture recognition using spherical anglefeatures and hidde Markov models, Adv. Hum. Comput. Interact. 2015 (2015) 7, http://dx.doi.org/10.1155/2015/785349 (2015).

[12] Archana Ghotkar, Pujashree Vidap, Kshitish Deo, Dynamic hand gesture recognitionusing hidden Markov model by Microsoft kinect sensor, Int. J. Comput. Appl. 150

[13] (2016), http://dx.doi.org/10.5120/ijca2016911498.[35] Pablo Barros, Nestor T. Maciel-Junior, Bruno J.T. Fernandes, Byron L.D. Bezerra,Sergio M.M. Fernandes, A dynamic gesture recognition and prediction system usingthe convexity approach, Comput. Vis. Image Understanding 155 (2017) 39–149.

[14] Sachin A. Urabinahatti, Apeksha S. Bandekar, Human-computer interaction usinghand gestures, Int. J. Recent Innov. Trends Comput. Commun. 2 (5) (2014)1305–1311.

[15] Gholam Reza Amayeh, Ali Erol, George Bebis, Mircea Nocilescu, Accurate andEfficient Computation of High Order Zernike Moments, ISVC 2005, Springer-Verlag,Berlin, 2005, pp. 462–469.

[16] Godlin Atlas.L,Kumar Parasuraman. "Detection of retinal hemorrhage from fundus images using ANFIS classifier and MRG segmentation" Biomedical Research 2018; 29 (7): 1489-1497.

[17] Yogesh Kumaran, Chandrashekar M. Patil, "A Brief Review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques" International Journal of Recent Technology and Engineering, Vol.7 (4S2), 2018, pp. 310-320.

[18]ZeynabMohammadpoory, MahdaNasrolahzadeh, NaghmehMahmoodian,JavadHaddadnia, " Automatic identification of diabetic retinopathy stages by using fundus images and visibility graph method" MeasurementVolume 140, July 2019, Pages 133-141