Skin-based adaptive background subtraction for hand gesture segmentation

Conference Paper · December 2015

DOI: 10.1109/ICECS.2015.7440242

Mohammed Sharaf Sayed
Egypt-Japan University of Science and Technology

Mahmoud I. Abdalla

Skin-based Adaptive Background Subtraction for Hand Gesture Segmentation

Rania A. Elsayed[1], Mohammed S. Sayed[1,2], Mahmoud I. Abdalla[1]
[1]Electronics and Communications Engineering, Zagazig University, Zagazig, Egypt,
en_rania_helika@yahoo.com, msayed@zu.edu.eg, mabdalla@zu.edu.eg
[2]Electronics and Communications Engineering, Egypt-Japan University of Science and Technology, New Borg El-Arab, Egypt,
mohammed.sayed@ejust.edu.eg

***Abstract*—Hand detection and localization is one of the challenging problems in image processing. It is the major first step in many computer vision applications like hand gesture recognition system that often requires as input the location of the hands. It is followed by two other stages; tracking and recognition. The accuracy of the later two stages strongly depends on the quality of the first one. There are many challenges in hand gesture segmentation such as changing illumination, shadows, and complex backgrounds. Even though there are several approaches for detection of moving hand area, most of the existing algorithms produce false positives. This paper proposes a robust hand gesture segmentation method that is based on adaptive background subtraction with skin color based threshold. The proposed method aims to automatically segment the hand gesture from a given video under different illumination conditions and complex backgrounds. Experimental results show that the proposed method is accurate, robust, reliable, and significantly reduces false positives. Besides that, this method provides high detection rate compared to other commonly used method for hand gesture segmentation.**

Keywords—Hand gesture segmentation; Background Subtraction; Skin Color Detection; YCbCr Color Space

I. Introduction

With the ubiquity of new information technology and media, more effective methods for human computer interaction (HCI) are being developed which do not rely on traditional devices such as keyboards, mice, and displays. With the development of hand gesture recognition techniques, hand detection have become very important preprocessing steps for designing fully automatic recognition systems [1]. Hand segmentation is the major step for the analysis and identification of a gesture. Quality of the gesture segmentation has direct impact on accuracy of recognition systems. The goal of hand segmentation algorithms is to precisely locate all regions that contain human hands in colored images obtained with video cameras. The task of hand segmentation is still a challenging task in many gesture recognition systems. The main difficulties in hand gesture segmentation are illumination changes, shadows, and complex background [2], [3]. The algorithms used for hand segmentation can be categorized as skin color detection algorithms, moving object detection algorithms, and hybrid algorithms. Each group of algorithms has its own advantages and disadvantage.

The algorithms in the first category aim to determine whether the pixel color is the same as human skin or not. These methods face difficulties like different skin tones, scene illuminations, and the possibility of having background pixels with the same color as the human skin. The goal of the algorithms in the second category is extracting all moving parts precisely in a video sequence. This method is faced by many challenges such as changes in illumination and changes in the background that can be due to movement of tree leaves and shadows cast by moving objects. The third category is hybrid algorithms that combine both background subtraction and skin color detection such as the one presented in [4]. These techniques try to overcome the disadvantages of the first two categories and take the advantage of each category.

This paper proposes a new hybrid approach to overcome the disadvantages of existing techniques for hand gesture segmentation using skin color based threshold integrated in the background subtraction thresholding process. The rest of the paper is organized as follows: Section II discusses an overview of the existing methods used for hand segmentation in video stream. Section III presents the proposed algorithm. Section IV demonstrates the experimental result for different cases. Section V presents the results performance analysis. Section VI concludes this paper.

II. Related Work

In this section, we review number of hand segmentation algorithms. The three popular techniques that are mostly used in hand detection are discussed in the following subsections:

A. Skin Color Detection Algorithm

Skin color detection is one of the important research areas in many of computer vision applications such as, hand gesture recognition, face detection and tracking [5]. The skin color segmentation algorithms segment the pixels of an image into skin and non-skin regions based on the skin color of the object of interest in the input video. This is based on the fact that the color of the human hand across all races agrees closely in its chrominance value and varies mainly in its luminance value. A challenging problem is how to select color models that are suitable for skin pixel classifications under different varying conditions [6]. The RGB color space is the default color space for most available image formats. This model has a serious drawback that the R, G, and B components are highly correlated, thus RGB color space does not separate luminance and chrominance. For this reason, the RGB color space is not a very favorable choice for color-based recognition algorithms. As for YCrCb space, the RGB components are separated into luminance component(Y) and chrominance components (Cb,Cr), therefore, the luminance information can be removed easily which makes it insensitive to brightness. Therefore, this color space is popular in the skin color based segmentation. Advantages of processing images in the YCbCr space can be found in [6], [7]. To do skin segmentation, there is a need for building a decision rule that will discriminate between skin and non-skin pixels. The best result for detecting the skin pixels can be achieved by using the following rule:

$$\begin{gathered} 76 < Cb < 126 \\ 132 < Cr < 173 \end{gathered} \tag{1}$$

Based on these thresholds, we get a binary image showing skin and non-skin regions. These methods have essential problem: Skin segmentation in video with complex background that contains colors similar to skin can yield small isolated groups of pixels that may be mistakenly segmented as skin.

B. *Moving Object Detection*

Moving object detection is a major step in many high level computer vision applications. Background subtraction (BGS) is a famous moving object detection technique. The basic idea of background subtraction is to subtract from the current image a reference image that often called background image or background model, by using algorithm based on pixel comparison to extract all moving objects inside of the video frame, and to remove the stationary components (i.e. background). BGS works well when the camera is static [8]. The equation of the BGS algorithm is as follows:

$$D_{foreground}(x,y) = \left| I_{current}(x,y) - B_{background}(x,y) \right| \tag{2}$$

where $I_{current}(x,y)$ is the current frame, which has moving objects, $B_{background}(x,y)$ is the background image that represent the background part of the scene, which has no moving objects, and $D_{foreground}(x,y)$ is the difference image between current frame and background image that contain all moving objects. A threshold value *th* is then set to get a binary image $BW(x,y)$ that contains all the moving objects:

$$BW(x,y) = \begin{cases} 1 & if \ D_{foreground}(x,y,t) \geq th \\ 0 & if \ D_{foreground}(x,y,t) < th \end{cases} \tag{3}$$

where *th* is a pre-defined threshold. When a value of a pixel in the difference image exceeds a given threshold value *th*, the pixel is considered as an object or foreground pixel; otherwise as a background pixel [9]. Selecting suitable threshold is tricky and the background performance is severely affected by it. In addition, the threshold has to be adaptive or it must be changed based on the video sequence. The BGS approach overcomes the problem of stationary background objects that have a color distribution similar to human skin.

The most prominent disadvantages of this method are: 1) Hands have to be the only moving objects in the field of view to be segmented. 2) Performance of BGS algorithms is severely affected by shadows and dynamic background (e.g. leaves movement. 3) Furthermore, in these methods, it is difficult to find global threshold to achieve accurate segmentation of all moving objects.

C. *Hybrid Techniques*

To distinguish moving skin objects from stationary background, we can use a combination of previously mentioned techniques to overcome the disadvantages of each technique when implemented individually and exploit the advantages of each one. In [4], Karishma, and Lathasree integrated skin color detection and background subtraction (BGS) algorithms by using a logic "AND" operation to suppress the false positives and improve hand detection performance. The main drawbacks of their method are:

- Traditional background subtraction algorithms were based on a unique threshold value arbitrarily decided. However, when processing images of different people taken in different imaging conditions, the use of only one single threshold value is not adapted to deal with the wide range of variations.
- It is required to have continuous background modeling and updating to adapt for varying illumination, light effects, background structures, and background change from whether change and repetitive motion from clutter.

III. The Proposed Algorithm

In order to overcome these disadvantages of hand detection algorithms, we propose a new method, which is a combination of background subtraction to detect moving hand area precisely in a real time video sequence using a threshold based on skin color values to improve the segmentation process. To detect the color of the skin, it is important to select first the right color space to be used [6]. In our proposed algorithm, we convert video sequences to YCbCr format. The proposed algorithm can be divided into the following steps:

A. *Pre-processing Stage*

Noise removal pre-processing step is very much required task in most video processing systems. Median filter is applied as a pre-processing step on the acquired video frames to remove noise and improve the video.

B. *Segmentation*

Since our target is to segment moving hands, we started with the basic BGS algorithm to separate the moving objects. Then BGS was modified such that it detects only moving hands. The basic background subtraction algorithm in the YCbCr domain can be represented with the following equation:

$$D(x,y,t)_{in\ ycbcr} = \left| I(x,y,t)_{in\ ycbcr} - B(x,y,t)_{in\ ycbcr} \right| \tag{4}$$

where $I(x,y,t)_{in\ ycbcr}$ is the current frame at time *t*, which has moving objects, whether hands, shadows or moving leaves, $B(x,y,t)_{in\ ycbcr}$ is the background image at time t, which has no moving objects and initially it is the first frame in the video sequence, and $D(x,y,t)_{in\ ycbcr}$ is the difference image between current frame and background image at time *t*. Therefore, this difference frame contains the moving objects found in the current frame.

Since skin color is located within a specific range of color that is usually different from the color of other moving objects in video, then using skin based threshold we can adjust the BGS method to segment moving hands only. To distinguish hands, the skin range of color in Cb and Cr are used to threshold the difference image $D(x, y, t)_{in\ ycbcr}$. Thresholding values of Cb and Cr are selected empirically from nearly 20 image sequences. The best result was found by using the following values for detecting moving skin pixels:

$$8 < Cb < 23$$
$$11 < Cr < 22 \tag{5}$$

Based on the values of the Cb and Cr thresholding, we get a binary image corresponds to only moving objects that have skin color and these objects are hand gesture and face if appears in image. This binary image contains all of the essential information about the position and shape of the foreground moving objects with skin color. Hence, there is a need for further elimination of other objects with skin colors such human face [8].

C. Post-Processing with Morphological Operators

To further refine the initial segmentation results, a series of morphological operations are applied. To get the output match with the required results, the steps followed are:

- Closing: To fill up gaps of the objects in binary mask to confirm the consistency in objects.
- Opening: To eliminate unwanted small objects while preserving the shape and size of larger objects in image.

D. Hand Localization of Segmented Image

Thresholding of color is not enough to detect and localize human hands in video. If the output segmented image contains any other moving parts of the body that have the skin color such as human face then it will be detected by the system as skin region [8]. Therefore, we use human face features method to filter out face object and detect only hand region in video/image [5]. Then, we identify the location of each hand through plotting rectangular box around these highly probable human hands in original video.

The proposed algorithm has major advantages. It provides the most complete feature data for hand object, and has a high hand detection rate when compared with traditional methods. Besides that it does not need to update the background model to adapt for the monitoring scene changes since it depends on detecting only skin moving objects and this is a power point of the proposed algorithm. And thus it works strongly under varying lighting conditions, shadows, and complex backgrounds.

IV. Experimental Results And Discussion

In this section, the experimental results of the proposed method are compared with skin color-based method [7], BGS method [8], and hybrid method in [4]. These methods were applied on different video sequences, which have been either obtained from internet or captured using fixed camera under different illumination conditions and different backgrounds.

From Fig. 1, the results of all methods are fairly good in case of simple background except BGS method; the whole arm will appear since it is moving as well as the hand. In Fig. 2, the results of the skin method are very bad because any pixel with skin color appears. Also results of BGS method are very bad, because all the moving objects appear as in Fig. 2 and Fig. 3. Although the results of the hybrid method are better than the two preceding methods, it was affected by the bad segmentation of the BGS algorithm which is ANDed with the skin color results. Hence a large part of the hand object is missed. But in our method, the results are very good as skin regions are detected with high accuracy and are not affected by shadow and leaf movement; because the skin color threshold is integrated in the BGS method to force segmenting moving objects with skin color from the beginning since neither shadow nor tree leafs have the chrominance range of the human skin. Thus, we overcome the drawbacks of the other methods.

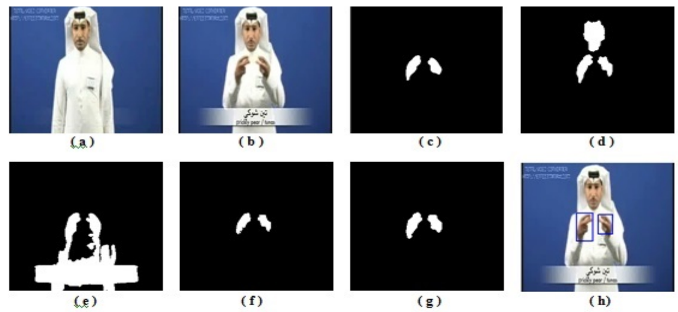

Fig. 1. Results for normal background. (a) Background image. (b) Current frame. (c) Ground truth. (d) Skin color results. (e) BGS results. (f) Hybrid results. (g) Proposed method results. (h) Bounding box for the proposed method.

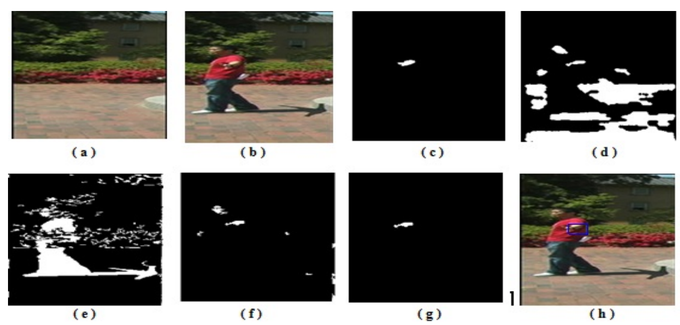

Fig. 2. Results for background containing skin colored constant objects, shadows and leaves movement. (a) Background image. (b) Current frame. (c) Ground truth. (d) Skin color results. (e) BGS results. (f) Hybrid results. (g) Proposed method results. (h) Bounding box for the proposed method.

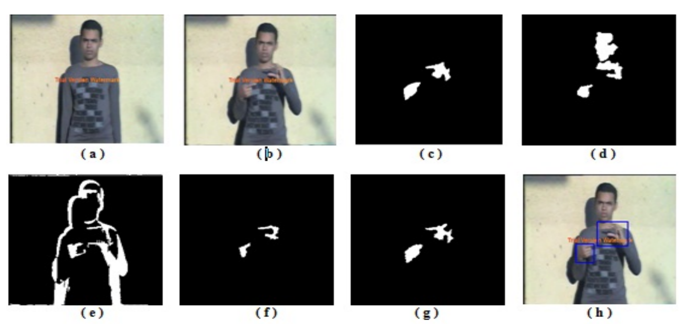

Fig. 3. Results for background containing human shadow. (a) Background image. (b) Current frame. (c) Ground truth. (d) Skin color results. (e) BGS results. (f) Hybrid results. (g) Proposed method results. (h) Bounding box for the proposed method.

The performance of the proposed method was also evaluated quantitatively by comparing the resultant segmented images with ground-truth images pixel by pixel. A ground-truth dataset for each video was manually created by labeling each frame with the two classes, hand and non-hand regions. The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values were found with respect to the Ground Truth images. Based on these values, the performance of the proposed algorithm is assessed according to eight distinct metrics for each image sequence. These metrics are given as [9], [10]:

1. Recall (Re) = TP/ (TP + FN)
2. Precision (Pr) =TP/ (TP+ FP)
3. Specificity (Sp) = TN/ (TN + FP)
4. False Positive Rate (FPR) =FP/ (FP + TN)
5. False Negative Rate (FNR) =FN/ (FN + TP)
6. Accuracy (Ac) = (TP + TN)/ (TP + FN + FP + TN)
7. Percentage of Wrong Classifications (PWC) = (FN + FP)/ (TP + FN + FP + TN)
8. F-measure =2(Pr*Re)/ (Pr + Re)

Recall, Precision, Specificity, Accuracy and F-measure are in the range of 0 and 1. In the ideal case, their values should be 1. FPR, FNR, and PWC are in the range of 0 and 1 and their ideal values should be 0. The performance results of the proposed method using these methods are given in Table I. The average values over six tested sequences are shown in Fig. 4 and Fig. 5. It can be noted from Table I that the proposed method works well regardless of the lighting conditions and/or complex backgrounds. The results of the proposed method are more accurate and have the highest performance (i.e. highest F-score). Also, from Fig. 4, the proposed algorithm has high Recall, Precision, Specificity and F-measure values, as well as has low FPR, FNR, and PWC values as shown in Fig. 5.

TABLE I. THE PERFORMANCE ANALYSIS FOR ALL METHODS

Videos	*Algorithm*	*Recall*	*Precision*	*FPR*	*PWC*	*F-Measure*
Video 1 shown in Fig. (1)	Skin Detection	0.9455	0.3696	0.023	0.0287	0.5315
	BGS Method	0.9714	0.0974	0.1578	0.1556	0.1770
	Hybrid Method [4]	0.8036	0.9667	0.0005	0.0039	0.8776
	Our Algorithm	0.9812	0.8546	0.0029	0.0032	0.9135
Video 2 shown in Fig. (2)	Skin Detection	0.7770	0.0120	0.2660	0.2658	0.0236
	BGS Method	0.8364	0.0304	0.1110	0.1112	0.0586
	Hybrid Method [4]	0.3978	0.2623	0.0046	0.0071	0.3161
	Our Algorithm	0.8513	0.8609	0.0006	0.0012	0.8561
Video 3 shown in Fig. (3)	Skin Detection	0.7584	0.3120	0.0328	0.0368	0.4422
	BGS Method	0.4488	0.0743	0.1096	0.1180	0.1275
	Hybrid Method [4]	0.4352	0.8977	0.0010	0.0118	0.5862
	Our Algorithm	0.8921	0.9006	0.0019	0.0040	0.8963

V. CONCLUSION

This paper presents a hand segmentation method to automatically segment the hand gesture from a given video under different luminance conditions and various complex backgrounds. The proposed method can effectively extract a hand even if there are skin-colored objects in background or there are other moving objects such as moving tree leaves and shadows. Meanwhile, using threshold for Cb and Cr components makes the proposed method invariant to varying illuminations, so the proposed method performs hand segmentation robustly under dynamic lighting conditions. The result section shows that the proposed approach is accurate and robust.

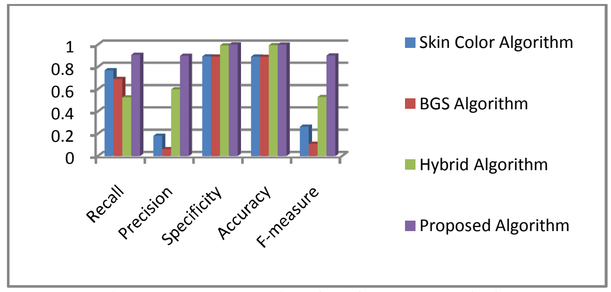

Fig. 4. Average results of performance analysis.

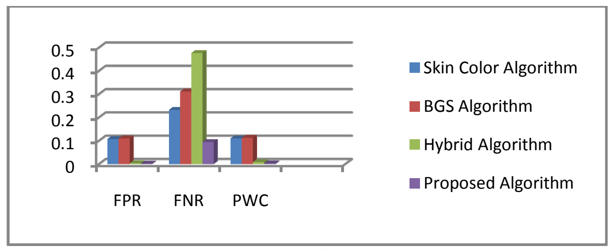

Fig. 5. Average results of performance analysis.

REFERENCES

[1] M. Segundo, L. Silva, P. Bellon, C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," IEEE Transactions On Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 40, No. 5, pp. 1319-1330, Oct. 2010.

[2] S. S. Rautaray, A. Agrawal, "Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey," In Artificial Intelligence Review (AIR), Springer Netherlands, vol. 43, No. 1, pp.1-54, January 2015.

[3] A. S. Ghotkar , G. K. Kharate, "Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction," International Journal of Human Computer Interaction(IJHCI), Computer Science Journal, Malaysia, vol. 3, No. 1, pp. 15-25, April 2012.

[4] S. N. Karishma, V. Lathasree, "Fusion of Skin Color Detection and Background Subtraction for Hand Gesture Segmentation," International Journal of Advanced Trends in Computer Science and Engineering ICETETS, vol. 3, No.1, pp. 13 – 18, February 2014.

[5] H. Shehadeh, A. Khalaf, M. Hassaweneh, "Human Face Detection Using Skin Color Information," Electro/Information Technology (EIT), IEEE International Conference on, May 2010.

[6] N.Krishnan, Ravi Subban, R.K.Selvakumar et al., "Skin Detection Using Color Pixel Classification with Application to Face Detection: A Comparative Study", IEEEInt. Conf. on Computational Intelligence and Multimedia Applications, vol. 3, pp. 436–441, 13-15 Dec. 2007.

[7] D. N. Chandrappa, M. Ravishankar, D. R. RameshBabu, "Face detection in color images using skin color model algorithm based on skin color information," 3rd International Conference on Electronics Computer Technology(ICECT), IEEE, vol. 1, pp. 254–258, April 2011.

[8] A. D. Gotmare, M. Sambath, S. Ravi, "Tracking And Recognizing The Moving Object In Real Time Using Frame Difference Method," in the International Conference on Advanced Engineering & Technology, Chennai, March 2012.

[9] M. Awney, M. S. Sayed, F. E. Abd El-Samie, "Hybrid Moving Object Detection Algorithm with Controlled Temporal Scale Frame Difference," IEEE, ICCTA 2013, pp. 29-31 October 2013.

[10] A. Kumar, C. Sureshkumar, "Background subtraction based on threshold detection using modified K-means algorithm", IEEE Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, pp. 378-382, 21-22 February 2013.