

A Hybrid Hand Detection Algorithm for Human Computer Interaction using Skin Color and Motion Cues

Komal Sheth
 Computer Engineering
 Department, SCOE
 Pune University

Nitish Gadgil
 NTPC Ltd
 Mouda
 Nagpur

P.R. Futane
 HOD, Computer Engineering
 Department, SCOE,
 Pune University

ABSTRACT

For complex background, a hybrid hand detection algorithm based on the hand motion and the skin color is introduced. Gaussian Mixture Model (GMM) is used for the Foreground Detection. Skin Detection is performed on Foreground objects. For the purpose of Skin Detection a Look-up Table (LUT) is generated by collecting large number of skin color pixels in YCbCr color space. Skin segmentation is performed by thresholding each frame to obtain a binary image of the skin color area. Morphological operations are used to enhance the segmentation results. Experimental results show that the proposed algorithm can accurately detect different hand shapes under complex background, which is the most necessary step in the HCI.

Keywords

Skin Color Segmentation, Computer Vision, hand Detection, Foreground detection.

1. INTRODUCTION

Accurate hand detection is the core of any system involving HCI. This paper mainly focuses on the hand detection. Hands are the most widely used body parts during gesturing.[22]. This statistics is shown in Fig. 1 in the form of a pie chart. Skin color, motion or texture cues of the hand can be effectively used for hand detection. Using skin color cue has some advantages; it allows faster processing and is robust to resolution changes. But this technique is plagued by some factors such as Illumination, Camera characteristics, Ethnicity. [20]. The techniques used for the skin color representation and segmentation are given in Fig.2.

This paper focuses on the pixel based skin segmentation techniques. As the hand moves during gesturing motion cue of the hand is also helpful in hand detection. In this case the moving hand in a frame becomes foreground and the rest remains background.

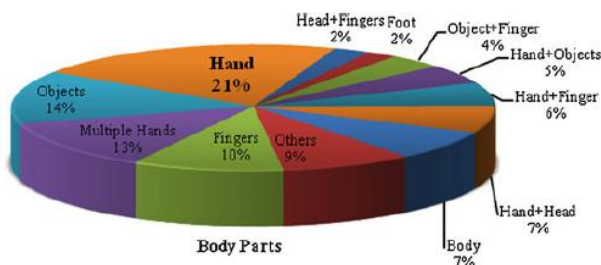


Fig 1: Pie chart showing different body parts used for gesturing in the literature [22]

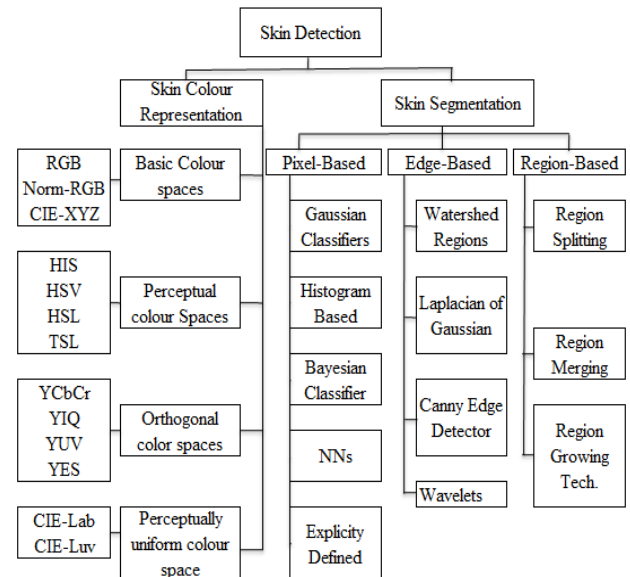





















Fig 2: Classification of the major aspects of skin detection [15]

2. PREVIOUS WORK

In the literature, skin color cue is extensively used for the purpose of hand detection. Table 1 provide the results of the various skin segmentation techniques used previously, but only skin color cue is used for the segmentation and background is not much complex in most of the cases. The pixel based skin segmentation technique; explicitly defined Skin region makes use of the fact that the human skin forms a small cluster when represented in color space. This method makes use of the explicitly declared decision boundaries for the components of the color space in order to detect the skin. YCbCr color space was used in [1]. The skin color cluster is very compact in the YCbCr space [1]. $100 \leq C_b \leq 140$ and $140 \leq C_r \leq 165$ these thresholds were applied to detect the skin pixels. The threshold gave very promising results with the ORL database. Similar technique was employed in [2] but with slightly different thresholds $C_b=[133,173]$ and $C_r=[77,127]$. The skin detection was done in YCbCr as it has a great adaptability to video coding hence reducing the computational time. These values were found suitable for the database used. Representation of the skin color is affected considerably with variation of illumination, ethnicity, camera used for capturing the images etc. Skin colors of individuals cluster differently in the color space [3]. In order to overcome this challenge a method using combination of three color spaces involving RGB, HSV and YCbCr for the

Table 1. Summary of the Skin Segmentation Methods [15]

I: Indoor; VI: Varying Illumination; CB: Complex Background; O/B-LUT: Ordinary/Bayesian Look Up Table

Work	Input	Output	Color space	Method	Expt. Cond	Data-Base
[16]			HSV	Explicit limits	I +VI+ CB	Own
[3]			YCb Cr	Free form model	I+VI +CB	Own
[8]			RGB	O-LUT	I + CB	Own
[9]			Five Color space	B-LUT	I+O+ CB	ECU Face Data Base
[6]			HSV	O-LUT	I+O	Own
[17]			YCb Cr	GMM	I +O +VI	Own
[18]			YCb Cr	GMM	-	Own
[19]			RGB +YC bCr+ HSV	Explicit Limits	I+O	Own
[23]			RGB +HS V+Y CbCr	Neural Network	I +VI	Own
[1]			YCb Cr	Explicit Limits	I + VI	ORL

purpose of skin segmentation is developed [4]. The thresholds of each color space were provided. Skin likely-hood binary image is obtained for each color space independently and at the end all the likely-hood images are combined to form the final segmented skin patch. The method proved to be efficient with the correct classification rate of 95.18% with complex background. The accuracy is found to improve at the cost of computational time. However, the threshold values change with color space. Performance of this method is very sensitive to illumination variance also it gives high true positive rate at expense of high false positive rate.

Non-parametric modelling techniques like histogram based method is used in [6], HS color space is used for the skin representation. The face detector could detect faces with the accuracy of 88.9%. The result concludes that the Look-Up Table (LUT) approach of skin modelling is computationally efficient. This approach was also adopted by [5], [7]. Bayesian method is used in [9], [8], [10], [7]. However, Non-parametric models lack the ability to interpolate the skin distribution, require large training set for better accuracy. The

performance of the model is affected by the degree of skin and non-skin overlap also these methods require high storage.

Parametric modelling techniques include, Single Gaussian Modelling (SGM) and Gaussian Mixture Modelling (GMM). SGM approach is used in [11] for skin color modelling using YCbCr space for her own image and video database. The performance of the SGM with GMM is compared in [14] by collecting 800 images containing people of different ethnic groups and concluded that the GMM is better than SGM. GMM with 5 components is used in [12] as a generic model to classify skin pixels. Then for further improvement of the accuracy the Support Vector Machine (SVM) classifier is used to predict the correct Gaussian for the classification. The false positive rate reduced by 8% using the two step process. A very comprehensive survey covering more than 100 publications of foreground and background modelling techniques using GMM is presented by [13]. However, these methods are computationally expensive because of the conventional EM algorithm used for parameter estimation.

Motion cue of the hand is not very popular. Also very little work has been done by using combination of two different cues. This paper exploits the possibility of obtaining a better hand detection results by combining motion cue and skin color cue in a hybrid approach.

3. PROPOSED APPROACH

Several methods of skin segmentation mentioned in the earlier section have certain limitations. Our goal is to detect hand in complex background as well as in varying illumination conditions. In order to overcome the imitations of using single skin segmentation method for hand detection; a hybrid approach using motion of the hand and faster skin segmentation is presented. Fig. 3 provides the basic flow of the algorithm and the techniques used at each stage.

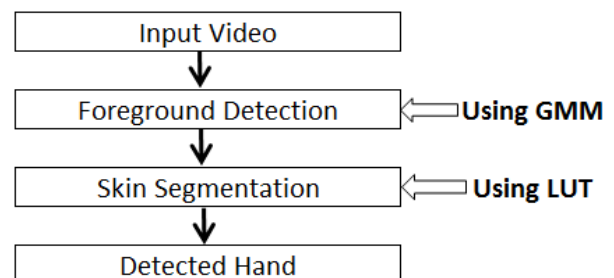


Fig 3: Proposed Hybrid approach

3.1 Foreground Detection

This is first stage of the proposed approach. This stage involves use of Gaussian Mixture Modelling (GMM) for foreground detection. The number of Gaussians to be considered depends upon the multimodality of the background available memory and computational power [13]. It was proposed by [21] to set number of Gaussians from 3 to 5. The principle behind the foreground detection is given below. Fig. 4 shows the initial steps of the foreground detection algorithm.

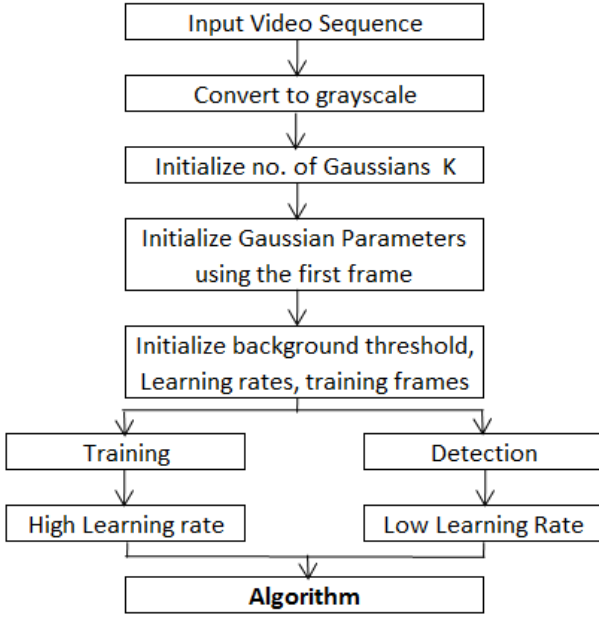


Fig 4: Initial Steps of Foreground Detection algorithm

The algorithm block shown in fig. 4 decides whether a pixel belongs to foreground or background. Initially the Gaussians are arranged in descending order of their weights. This is done assuming the Gaussian with the highest weight represents background due to the fact that background is more present than the foreground objects. [13] Initially a decision is to be made regarding how many Gaussians represent background for a particular pixel. In order to find that equation (1) is used,

$$B = \underset{B}{\operatorname{argmin}}_K \left(\sum_{i=1}^K \omega_{i,t} > BG_THR \right) \quad (1)$$

$\omega_{i,t}$ represents weight of the i^{th} Gaussian at time t .

BG_THR represents background threshold.

B gives the number of Gaussians that represent background and the rest represent foreground.

Now, pixel under consideration belongs to which Gaussian is determined as follows, when a new frame incomes at time $t+1$,

$$d = \mu_t - \text{gray value of pixel}_{t+1} \quad (2)$$

A match test is carried out for each Gaussian as given in equation (3),

$$d^2 < m\sigma_{i,t}^2 \quad (3)$$

Where m is constant and $\sigma_{i,t}^2$ is the variance of the i^{th} Gaussian at time t . here two cases can occur as follows,

Case 1: A match is found with one of the K Gaussians, then $\text{match}=1$. If that Gaussian is representing background, according to the results of the equation (1), then the pixel is classified as background else it is foreground. The Gaussian parameters are updated as follows,

$$\omega_{i,t+1} = \omega_{i,t} \times \alpha + \alpha \times \text{match} \quad (4)$$

$$\mu_{i,t+1} = \mu_{i,t} - \text{match} \times (\alpha \times d) \quad (5)$$

$$\sigma_{i,t+1}^2 = \sigma_{i,t}^2 + \text{match} \times (\alpha \times (d^2 - \sigma_{i,t}^2)) \quad (6)$$

The mean and variance of the unmatched Gaussians are kept same and only their weights are updated as follows,

$$\omega_{i,t+1} = \omega_{i,t} \times (1 - \alpha) \quad (7)$$

Case 2: No match is found with any of the Gaussians, the pixel is considered as foreground. The Gaussian with the least weight (1) is replaced by a new Gaussian as follows,

$$\omega_{l,t+1} = \text{Low Prior weight} \quad (8)$$

$$\mu_{l,t+1} = \text{Current Pixel Gray value} \quad (9)$$

$$\sigma_{l,t+1}^2 = \text{Large Initial variance} \quad (10)$$

Again the Gaussians are arranged in descending order of their weight and the process is repeated.

3.2 Skin Detection

Next stage of the hybrid approach is applying skin segmentation on the foreground objects in order to detect hand. Non-Parametric skin modelling approach is used for skin segmentation. Flow chart of the Look-Up Table generation and skin detection is shown in fig.5. This algorithm gives a final binary image of the hand in each frame of the input video sequence. Then morphological operations are carried out on the final binary image so as to get accurate hand detection.

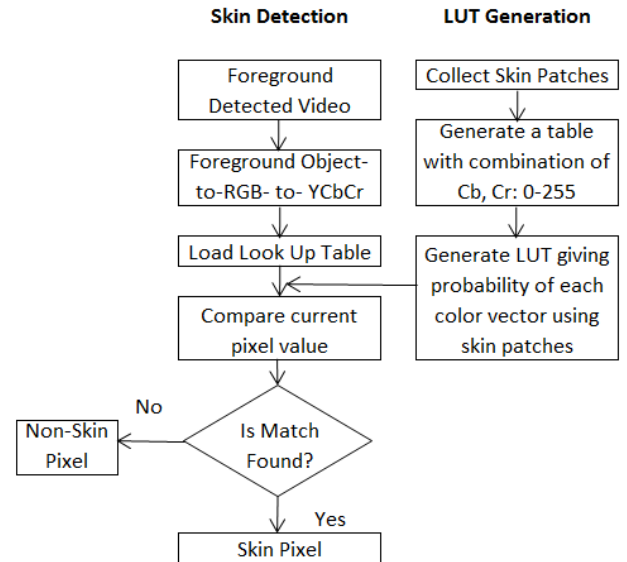


Fig 5: Flow chart of Skin Detection

4. RESULTS AND DISCUSSION

The difference between Foreground Detection stage of the proposed algorithm and the one given by [21] is that, in this case few initial frames of the video sequence are used for the online background learning and hand is not inserted in front of the camera till the background learning is over. This is done in order to learn the background efficiently and it gives more accurate foreground detection results. Table 2 gives the details of the parameters used for Foreground Detection.

For Skin Detection Number of skin patches of people from various ethnic groups are collected under different illumination conditions. Fig.6 shows some of the skin patches collected. These patches are in RGB color space. These are converted to YCbCr color space for generating LUT. YCbCr color space is adopted by analyzing the distinct advantages it offers in skin detection, as mentioned in [3].

Table 2. Parameters used for Foreground Detection

Parameter	Value
K	2
BG_THR	0.7
Training Frames	75
α (Training/Detection)	0.9/0.003
m	2.5



Fig 6: Some of the collected skin patches

A threshold for classifying a pixel as skin pixel or non-skin pixel is set by trial and error basis. Fig. 7 shows the output of the proposed hybrid approach in complex background, input frame, Foreground Detected image and the detected hand as a binary image. The complex background shown in Fig.7 (a) is filtered out to get the Foreground, fig. 7 (b) and after application of skin detection algorithm we finally get the detected hand in that frame, fig. 7 (c).

Fig. 8 shows the comparison of the proposed hybrid approach with other approaches. It can be easily interpreted from the fig. 8 that the proposed approach outperforms the other approaches considerably. Using only LUT approach or explicitly defined skin region approach increases the chances of false skin classification in complex backgrounds. The proposed method shows promising capabilities of handling complex backgrounds.

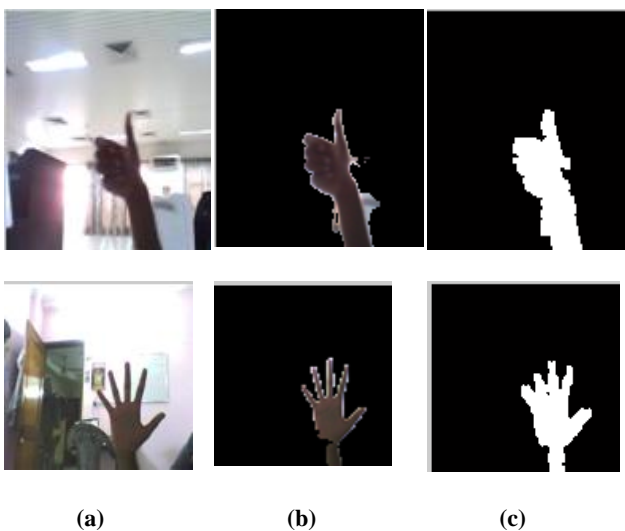


Fig 7: Hybrid approach, (a) Input Image, (b) Foreground, (c) Detected hand



Input Image

Method	Output	RGB image of Output
Only LUT		
Explicitly Defined Skin Region		
Proposed Hybrid Approach		

Fig 8: Comparison of Proposed approach with different approaches.

5. CONCLUSION

In this paper a hybrid hand detection algorithm is presented. The results show that human skin color can be a reliable cue for detecting hand in complex background, if combined with the motion cue. The motion cue is obtained using GMM and the skin color cue is obtained using LUT. In case of hand detection in complex background, the proposed algorithm outperforms the existing skin detection methods.

The hand detection in complex background is an initial step for every HCI systems and applications, for example, Sign Language Recognition. This work could be a foundation for the research work in Sign Language Recognition.

6. ACKNOWLEDGMENTS

The authors wish to thank to Mr. Shirin Dora for providing research papers as and when needed, also special thanks to Mr. Abhishek Badki for providing valuable suggestions throughout the research work.

7. REFERENCES

- [1] R. Swathi, "Composite Feature-Based Face Detection Using Skin Color Modeling and SVM Classification", Master's Thesis, Computer Science, Utah State University, 2012.
- [2] A. Garg; "Converting American Sign Language To Voice Using RBFNN"; Master's Thesis, Computer Science, Faculty of San Diego State University, Summer 2012.
- [3] A. Dawod, J. Abdullah and Md.J. Alam, "A New Method for Hand Segmentation Using Free-Form Skin Color Model", 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE), 2010.
- [4] S. Singh, D. S. Chauhan, M. Vatsa and R. Singh, "A Robust Skin Color Based Face Detection Algorithm", Tamkang Journal of Science and Engineering, Vol. 6, No. 4, pp. 227-234, 2003.
- [5] R. Porle, A. Chekima, F. Wong, and G. Sainarayanan, "Performance of Histogram-Based Skin Colour Segmentation for Arms Detection in Human Motion Analysis Application", International Journal of Electrical and Computer Engineering 4:15, 2009.
- [6] Q. Liu, G. Peng, "A Robust Skin Color Based Face Detection Algorithm", 2nd International Asia Conference on Informatics in Control, Automation and Robotics, 2010.
- [7] B. Zarit, B. Super and F. Quek, "Comparison of Five Color Models in Skin Pixel Classification", Proceedings. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999.
- [8] L. Howe, F. Wong, A. Chekima, "Comparison of Hand Segmentation Methodologies for Hand Gesture Recognition", Information Technology, ITSIM 2008. International Symposium, Volume:2, 2008.
- [9] S. Phung, A. Bouzerdoum, and D. Chai, "Skin Segmentation Using Color and Edge Information", Proceedings Seventh International Symposium on Signal Processing and Its Applications, vol:1, 2003.
- [10] Vezhnevets, V. and Andreeva, A. 2005. A Comparative Assessment of Pixel-based Skin Detection Methods, Technical report, Graphics and Media Laboratory, Moscow State University.
- [11] J. Rekha, J. Bhattacharya and S. Majumder, "Shape. Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition", 3rd International Conference on trends in Information Sciences and Computing (TISC), 2011.
- [12] Q. Zhu, K. Ting, C. Wu and Y. Wu, "Adaptive Learning of an Accurate Skin-Color Model", Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.
- [13] T. Bouwmans, F. El Baf and B. Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey"; Recent Patents on Computer Science Volume 1, No 3, pages 219-237, November 2008.
- [14] T.S. Caetano, D.A.C. Barone; "A probabilistic model for the human skin-color"; ICIAP01, pp. 279-283, 2001.
- [15] K. Sheth, N. Gadgil and P.R. Futane, "A Survey of Gesture Recognition Systems for Indian Sign Language Recognition", International Journal of Engineering and Advanced Technology (IJEAT), Vol.2, Issue.6, August 2013.
- [16] X. Li; "Vision Based Gesture Recognition System With High Accuracy"; Department of Computer Science, The University of Tennessee, Knoxville, TN 37996-3450, 2005.
- [17] R. Hassanpour, A. Shahbahrami and S. Wong; "Adaptive Gaussian Mixture Model for Skin Color Segmentation"; Proceedings Of World Academy Of Science, Engineering And Technology, Issn 1307-6884, Vol: 31, July 2008.
- [18] S. Zhao, W. Tan, S. Wen, and Y. Liu; "An Improved Algorithm of Hand Gesture Recognition under Intricate Background"; the First International Conference on Intelligent Robotics and Applications (ICIRA 2008); Part I. Springer-Verlag Berlin Heidelberg, 2008, pp. 786-794.
- [19] P. Patil, Y. Patil; "Robust Skin Colour Detection And Tracking Algorithm"; International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1 Issue 8, October – 2012.
- [20] P. Kakumanu, S. Makrogiannis, N. Bourbakis; "A survey of skin-color modeling and detection methods"; Elsevier, The journal of the pattern recognition society, 40 ,1106 – 1122, 2007.
- [21] C. Stauffer and W. E L. Grimson. "Adaptive background mixture models for real-time tracking." Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.. Vol. 2. IEEE, 1999.
- [22] Karam, M. 2006. A framework for research and design of gesture-based human computer interactions. PhD Thesis, University of Southampton.
- [23] A. Chaudhary and A. Gupta; "Automated Switching System for Skin Pixel Segmentation in Varied Lighting"; 19th International Conference on Mechatronics and Machine Vision in Practice (M2VIP12), Nov 2012.