

Research Article

Skin Segmentation using Ensemble Technique

S.M. Jaisakthi and S. Mohanavalli

SSN College of Engineering, Chennai, Tamil Nadu-603110, India

Abstract: Localizing potential skin regions in a color image forms a significant step in applications like face detection, face recognition, face verification, face tracking, gesture analysis, content-based image retrieval and human computer interaction. In this study, we present a pixel based skin segmentation algorithm with ensemble approach using Gaussian Mixture Model (GMM) classifier. Skin color features are extracted using RGB, HSV, YCbCr and CIE Lab color spaces and ensembled into a single feature vector which is used to train the GMM classifier. Comprehensive experiments have been conducted using three different datasets, SFA Database, ECU Skin Database and UCI Skin database. The skin detection rate of our proposed classifier is observed to be better than the existing works.

Keywords: Classifier ensemble, feature extraction, gaussian mixture model, skin segmentation

INTRODUCTION

Recently skin detection technique is gaining more importance since many applications like face detection, face recognition, face tracking, gesture analysis, content-based image retrieval and human computer interaction use skin detection as a preliminary step to localize skin regions. Kakumanu *et al.* (2007), discussed that detecting skin color is a challenging task because skin color in images are affected by illumination, background, camera modality and race. Skin color segmentation can be categorized as pixel based methods and region based methods. Pixel based method builds a skin classifier by explicitly defining a skin cluster boundary in some color space. Construction of rapid classifier is the main advantage of this method since it uses simple skin detection rules. Kovac *et al.* (2003) and Ahlberg, 1999, have concluded that high recognition rates can be achieved only by choosing a good color space along with the appropriate decision rules. Kruppa *et al.* (2002), Jedynek *et al.* (2003) and Yang *et al.* (1998), presented region based pixel methods that analyze the spatial arrangement of skin pixel during the detection stage.

Most of the existing works for skin detection is based on pixel based methods. In the past decade many research work has been reported on skin color pixel classification. Cho *et al.* (2001) proposed a adaptive skin color filter which is capable of adjusting the threshold values which effectively separates skin color regions. A dynamic skin color model was proposed by (Sun, 2010) that uses a local skin model to shift a globally trained skin model to adapt the final skin model. Cheddad *et al.* (2009) proposed a new color space which contains error signal derived by

differentiating the grayscale map and the non-red encoded grayscale version.

Phung *et al.* (2005) presented the issues pertaining to skin color pixel classification. The authors investigated eight different color representations, seven different levels of color quantization and nine different color pixel classification algorithms. A simple probabilistic model for classifying pixels as skin or non-skin pixel is proposed in Fkihi *et al.*, 2009. Rehg and Jones (2002) discussed a skin color detection algorithm based on self-adaptive skin color model which depends on the luminance Y.

Hassanpour *et al.* (2008) proposed an adaptive skin color segmentation method using Gaussian Mixture Model that can adapt the model parameters according to the changing imaging conditions such as lighting and noise. A self adaptive Gaussian mixture model for segmenting the foreground images from the background is proposed in (Chen and Ellis, 2014). The proposed method uses a dynamic learning rate with adaptation to global illumination to adjust the variation in illumination. Skin color segmentation using texture and k-means clustering was proposed by Pun and Ng, 2014 and used both color and texture features to improve the accuracy of skin detection.

The objective of this research work is to segment the potential skin regions in the given color images inspite of variation in race, background, illumination etc. In this study we propose a skin segmentation algorithm based on feature ensemble using various color spaces to improve the recall and accuracy rate. The skin detection rate of our classifier is much better than the existing works, Phung *et al.*, 2005 and Casati *et al.*, 2013.

LITERATURE REVIEW

Pixel based skin color segmentation algorithm classifies each pixel as skin or non-skin pixel based on thresholding technique. In this study, four different color space have been used for feature ensembling, which are discussed below.

RGB and normalized RGB color model: RGB color space is the most widely used color model to store and represent digital image. This model specifies the intensity of red, green and blue on a scale of 0 to 255. The settings of the three colors are converted to a single integer value by using the formula $RGB\ value = Red + (Green * 256) + (Blue * 256 * 256)$. RGB color model is not suitable for all applications since the red, green and blue color components are highly correlated.

The RGB color components are normalized in order to reduce the dependence of lighting. Normalized RGB can be obtained by simply normalizing the three color components which are given by:

$$r = \frac{R}{R+G+B}$$

$$g = \frac{G}{R+G+B}$$

$$b = \frac{B}{R+G+B}$$

The sum of these three normalized color components is 1 (r+b+g), the third component does not hold any significant information and can be dropped to reduce the space dimensionality. RGB color space is more popular among the researchers due to its simple transformation property and invariant to changes of surface orientation relatively to the light source when we ignore the ambient light source Brown *et al.*, 2001 and Skarbek *et al.*, 1994.

Hue Saturation (HS) color model: It will be more intuitive if a color can be described by the perceptual property of color such as Hue (H), Saturation (S) and Intensity (I). This is implemented via this color space where Hue defines the dominant color (such as red, green, purple and yellow) of an area; saturation measures the colorfulness of an area in proportion to its brightness. The intensity, lightness or value is related to the color luminance. In HSV color model values can be extracted using non-linear transformation of RGB color primaries. RGB to HSV transformation may be expressed as:

$$HSV = \begin{cases} H = \begin{cases} h, B \leq G \\ 2\pi - h, B \geq G \end{cases} \\ \text{where } h = \cos^{-1} \frac{1/2[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-G)(G-B)}} \\ s = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R, G, B)} \\ v = \max(R, G, B) \end{cases}$$

The transformation of RGB to HSV is invariant to high intensity at white lights, ambient light and surface orientations relative to the light source and hence, it can be a better choice for skin detection methods (Kakumanu *et al.*, 2007). The HSI color space (hue, saturation and intensity) attempts to produce a more intuitive representation of color. The I axis represents the luminance information. Converting colors from RGB to HSI is given by:

$$HSI = \begin{cases} H = \begin{cases} \theta, B \leq G \\ 360 - \theta, B > G \end{cases} \\ \text{where } \theta = \cos^{-1} \frac{1/2[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-G)(G-B)}} \\ s = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \\ I = \frac{1}{3}(R+G+B) \end{cases}$$

The main drawback of this color space is that hue is undefined if saturation is zero.

YCbCr color model: YCbCr color space has been defined in response to increasing demands for digital algorithms in handling video information and has become a widely used model in a digital video Boykov and Kolmogorov, 2001. In YCbCr color model color is represented as luminance (Y) computed as a weighted sum of RGB values and chrominance (Cb and Cr) computed by subtracting luminance from Red and Blue components:

$$Y = 0.299R + 0.587G + 0.114B$$

$$C_r = R - Y$$

$$C_b = B - Y$$

Due to its simple transformation and explicit separation of luminance and chrominance this color space can be a better choice for skin tone detection.

CIE Lab color model: CIE Lab is based on one channel for Luminance (L) and two color channels (a and b). It is the accurate mathematical model that emulates normal human color vision based on standard viewing conditions, light sources and a defined standard observer set by CIE. This color model is device independent. This color space is more suitable for digital manipulation than the RGB space Amanpreet and Kranthi, 2012. CIE Lab color model is directly based on CIE XYZ and the conversion from RGB to CIE Lab is given by:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.537150 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} X \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$L^x = \begin{cases} 116 * (Y / Yn)^{1/3} - 16 & Y / Yn > 0.008856 \\ 903.3 * Y / Yn & \text{otherwise} \end{cases}$$

$$a^x = 500 * (f(X / Xn) - (Y / Yn))$$

$$b^x = 200 * (f(Y / Yn) - (Z / Zn))$$

$$\text{where } f(t) = \begin{cases} t^{1/3} & t > 0.008856 \\ 7.787 * t + 16 / 116 & \text{otherwise} \end{cases}$$

Gaussian Mixture Model (GMM): A Gaussian Mixture Model (GMM) can be defined as a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs find application in Background Subtraction (BS), moving object detection, speaker recognition system, biometric and skin pixel detection. GMM parameters are estimated using iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from training data. The likelihood given a Gaussian distribution is given by:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma_i|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu_i)\right)$$

where,

d = The dimension of sample x

μ_i = The mean

Σ_i = The covariance matrix of the Gaussian

A Gaussian mixture model is a weighted sum of M Gaussian components which is given by:

$$p(x|\lambda) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i)$$

where, x is a d dimensional feature vector, $w_i, i = 1, \dots, M$, are the mixture weights with constraints $\sum_i w_i = 1$ and $g(x|\mu_i, \Sigma_i)$ where $i = 1, \dots, M$, are the Gaussian components.

MATERIALS AND METHODS

Proposed skin segmentation technique: Identifying skin pixels in an image is a vital step in most of the face processing applications. The traditional approach for skin segmentation extracts pixel values from a single color space in order to classify skin pixel. In this research work, we propose to ensemble the pixel values using RGB, HSV, YCbCr and CIE Lab color spaces into a single feature vector. The advantage of this ensemble approach is that the recall value is improved since the complimentary information are extracted using different color spaces.

Our proposed method extracts pixel values from the training set which contains only skin pixels. For each skin pixel, the pixel values are obtained using above mentioned color spaces and ensemble into a

single feature vector which contains complimentary information. In order to identify the number of ethnicity groups (M-Number of Gaussian mixture components) in the training set spectral clustering algorithm was applied since spectral clustering provides number of potential clusters in any given data set Ng *et al.*, 2001. The obtained mean and covariance from the spectral clustering is used as the initial parameters to model skin distributions as Gaussian mixtures.

In order to segment the pixels as skin pixels, log likelihood values are obtained using the Gaussian model which is given by:

$$p(x|skin) = \sum_{i=1}^M w_i g(x|\mu_i, \Sigma_i)$$

where, weight w_i , mean μ_i and covariance Σ_i are obtained by Expectation Maximization (EM) algorithm. A pixel is classified as skin pixel if the log likelihood value is above a threshold τ . The threshold value τ is fixed using Equal Error Rate (EER) method. Equal Error Rate (EER) is the value where false positives and false negatives are equal. It is also referred as crossover rate or Crossover Error Rate (CER). The lower the equal error rate value, the higher the accuracy of the GMM classifier. The threshold value τ is predetermined from the log likelihood value returned by the GMM classifier for the training samples. If the log likelihood value of a pixel falls below the threshold τ then it is labeled as skin pixel and non-skin pixel otherwise. The algorithm for skin segmentation using the ensembling technique is presented in Algorithm 1.

Algorithm 1 Algorithm for skin segmentation using Ensemble Technique:

- 1: Function SKIN SEGMENTATION
- 2: For all Images I in the training set Tr do
- 3: For all Pixels in an image I do
- 4: Extract pixels values for RGB, HSV, YCbCr and CIE Lab color spaces.
- 5: Form ensemble feature vector by combining the extracted pixel values into a single vector.
- 6: End for
- 7: End for
- 8: Apply spectral clustering in the ensemble feature space to find the number of mixture models M.
- 9: Build the GMM classifier with M mixture components using the obtained mean μ and covariance Σ .
- 10: Fix the threshold τ using the loglikelihood values obtained in the training phase.
- 11: For all Images I in the test set Tt do
- 12: For all Pixels in an image I do
- 13: If loglikelihood value $> \tau$ then
- 14: Label the pixel as skin pixel
- 15: Else
- 16: Label the pixel as non-skin pixel

17: End if
 18: End for
 19: End for
 20: End function

RESULTS AND DISCUSSION

Comprehensive experiments were conducted to evaluate our proposed ensemble technique to segment skin regions.

Data: To ascertain the efficiency our proposed work we have tested with three databases namely, SFA Database, ECU Skin Database (Phung *et al.*, 2005) and UCI Skin database. SFA database (Casati *et al.*, 2013) contains 3354 skin samples and 5590 non-skin samples. Both the skin and non-skin samples vary in dimension, from 1 pixel to 35×35 pixels. For training we have used 4,108,650 (35×35×3354) skin pixels. The database contains test set (original images) along with the ground truth. ECU skin database contains 4,000 color images from four races, with varying illumination and background along with the ground truth. From this database we have used 2500 images to train the Gaussian classifier and the remaining 1500 images for testing. UCI skin dataset (Bache and Lichman, 2013) contains RGB values of skin and non-skin pixels which were collected from the face images of FERET and PAL databases. The database includes 3 races (white, black and asian) with 50859 skin pixels and 194198 non-skin pixels. To train the GMM classifier with UCI database we have used 60% of the samples for training and the remaining 40% samples for testing.

Evaluation criteria: The performance of the proposed skin classifier is evaluated with accuracy (acc), precision (p) and recall (r). Accuracy is a measure to determine how well a classifier classifies a given sample as skin or non-skin pixel. It is calculated as total number of correct results out of total number of samples. Precision is calculated as the ratio of actual skin positives out of predicted positives and recall can be calculated as the ratio of actual positives out of positives. Accuracy, precision and recall can be calculated with the formula listed below:

$$\text{Accuracy (acc)} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total No. of Samples}}$$

$$\text{Precision (P)} = \frac{\text{True positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall (R)} = \frac{\text{True Positives}}{\text{Total No. of Positives}}$$

Performance of the proposed skin classifier: The performance of our proposed ensemble technique was

Table 1: Result of the proposed ensemble technique for SFA database

| | Precision | Recall | Accuracy |
|----------|-----------|--------|----------|
| RGB | 0.9367 | 0.9012 | 0.9478 |
| HSV | 0.9045 | 0.9356 | 0.9678 |
| YCbCr | 0.9148 | 0.9478 | 0.9636 |
| CIELab | 0.9445 | 0.9267 | 0.9564 |
| Ensemble | 0.9478 | 0.9563 | 0.9876 |

Table 2: Result of the proposed ensemble technique for UCI database

| | Precision | Recall | Accuracy |
|----------|-----------|--------|----------|
| RGB | 0.7291 | 0.8104 | 0.9698 |
| HSV | 0.6680 | 0.7678 | 0.9586 |
| YCbCr | 0.7554 | 0.9895 | 0.9735 |
| CIELab | 0.9805 | 0.6960 | 0.9748 |
| Ensemble | 0.9518 | 0.9776 | 0.9955 |

Table 3: Result of the proposed ensemble technique for ECU database

| | Precision | Recall | Accuracy |
|----------|-----------|--------|----------|
| RGB | 0.8986 | 0.8109 | 0.9414 |
| HSV | 0.8456 | 0.9177 | 0.9575 |
| YCbCr | 0.7521 | 0.8018 | 0.9632 |
| CIELab | 0.8802 | 0.7139 | 0.9641 |
| Ensemble | 0.8466 | 0.9879 | 0.9788 |

tested with individual color space using SFA Database, ECU Skin Database and UCI Skin database. In this study we have built five Gaussian mixture models (Four GMMs for four color spaces and one GMM for ensemble color space) to model the skin distribution in an image. To train the GMM we have used only the skin pixels from the database and for testing, the test images in the database were used. The output image produced by the classifier was compared pixel wise with the corresponding ground truth image. With the ensemble technique we have obtained a significant improvement in terms both the accuracy and recall value. Table 1 shows the results of the proposed ensemble technique for SFA database. For SFA database both HSV and YCbCr color space produced almost the same accuracy, but YCbCr produced high recall value when compared to HSV color space. By combining all the four color models we have obtained the recall value of 0.9563 and accuracy of 0.9876 which is higher than the individual methods.

We have also obtained improved results for both UCI and ECU skin databases. The results of the proposed ensemble classifier for skin segmentation using UCI and ECU skin databases are presented in Table 2 and 3, respectively. For UCI database our proposed ensemble technique produced the recall rate of 0.9776 and accuracy rate of 0.9955 which is higher than the individual color spaces. Similarly for ECU database we have obtained recall rate of 0.9879 and accuracy rate of 0.9788 which higher when compared to the individual methods.

For all the three databases we have obtained high recall rate with the ensemble technique. Since our proposed ensemble technique makes use of complimentary information we have obtained better recall value than the individual methods.

Comparison with the existing works: The results of the proposed work is compared with the methods

Table 4: Comparison of the proposed ensemble technique using ECU database

| Method | Accuracy |
|--|----------|
| Bayesian classifier (Phung <i>et al.</i> , 2005) | 89.79% |
| Multilayer perceptrons (Phung <i>et al.</i> , 2005) | 89.49% |
| Gaussian of skin using YCbCr (Phung <i>et al.</i> , 2005) | 82.67% |
| Gaussian of skin using CbCr (Phung <i>et al.</i> , 2005) | 85.57% |
| Gaussian of skin and non-skin using YCbCr (Phung <i>et al.</i> , 2005) | 88.92% |
| Gaussian of skin and non-skin using RGB (Phung <i>et al.</i> , 2005) | 85.76% |
| Proposed method | 97.88% |

Table 5: Comparison of the proposed ensemble technique using SFA and UCI database

| Method | Database | Accuracy |
|---|----------|----------|
| Multilayer (Casati <i>et al.</i> , 2013) | SFA | 92.71% |
| Multilayer Perceptron (Casati <i>et al.</i> , 2013) | UCI | 88.74% |
| Proposed method | SFA | 98.76% |
| Proposed method | UCI | 99.55% |

proposed by Phung *et al.*, 2005 and Casati *et al.*, 2013. The comparison of the proposed work with the existing work using ECU database is shown in Table 4. For ECU database we obtained improved performance in terms of accuracy. We have obtained 97.88% accuracy for this database which is 9% higher than the method proposed by Phung *et al.*, 2005.

We have also obtained better accuracy than the method proposed by Casati *et al.*, 2013 and the results are shown in Table 5. The proposed method produces 7% increase in accuracy for SFA database and 10% increase in accuracy for UCI database.

CONCLUSION AND RECOMMENDATIONS

In this research work we have proposed a novel method for segmenting the skin regions in a given image using ensemble technique. The proposed method extracts the skin pixels using four different color spaces namely RGB, HSV, YCbCr and CIELab and ensemble the extracted pixels into a single feature vector. The ensemble feature vector is then used to train the Gaussian classifier. Using the log likelihood values the thresholding technique classify every pixel in an image are classified as skin or non-skin pixels.

We have tested our new approach using three well known databases namely, SFA Database, ECU Skin Database and UCI Skin database. Out of these three databases we have obtained maximum recall rate of 0.9776 and accuracy rate of 0.9955 for UCI database. For all the three databases we have also obtained better recall rate than the individual color spaces. The proposed method also performs better than the methods proposed by Phung *et al.*, 2005 and Casati *et al.*, 2013. Our method produced 7% improvement in terms of accuracy when compared to the method proposed by Phung *et al.*, 2005. For the method proposed by Casati *et al.*, 2013 we obtained 6% improvement using SFA database and 9% improvement using UCI database.

In the future we intend to parallelize the proposed ensemble technique to reduce the processing time.

ACKNOWLEDGMENT

The authors would like to express their sincere thanks to the management of SSN College of Engineering, Chennai, India, for funding the High Performance Computing Lab (HPC Lab) where this research is being carried out.

REFERENCES

- Ahlberg, J., 1999. A system for face localization and facial feature extraction. Report No. LiTH-ISY-R-2172.
- Amanpreet, K. and B.V. Kranthi, 2012. Comparison between ycbcr color space and cielab colorspace for skin color segmentation. *Int. J. Appl. Inform. Syst.*, 3(4): 30-33.
- Bache, K. and M. Lichman, 2013. UCI Machine Learning Repository. Retrieved form: <http://archive.ics.uci.edu/ml>.
- Boykov, Y. and V. Kolmogorov, 2001. Energy minimization methods in computer vision and pattern recognition. *EMMCVPR 2001*, Sophia Antipolis France, September 3-5, 2001.
- Brown, D., I. Craw and J. Lewthwaite, 2001. A SOM based approach to skin detection with application in real time systems. In: Timothy F. Cootes and Christopher J. Taylor (Eds.), *Proceedings of the British Machine Conference*, pp: 51.1-51.10.
- Casati, J.P.B., D.R. Moraes and E.L.L. Rodrigues, 2013. Sfa: A human skin image database based on feret and AR facial images. *Proceeding of IX Workshop de Viso Computacional, Anais do VIII Workshop de Viso Computacional*.
- Cheddad, A., J. Condell, K. Curran and P. Kevitt, 2009. A skin tone detection algorithm for an adaptive approach to steganography. *Signal Process.*, 89(12): 2465-2478.
- Chen, Z. and T. Ellis, 2014. A self-adaptive Gaussian mixture model. *Comput. Vision Image Understanding*, 122: 35-46.
- Cho, K., J. Jang and K. Hong, 2001. Adaptive skin-color filter. *Pattern Recogn.*, 34(5): 1067-1073.
- Fkihi, S., M. Daoudi and D. Aboutajdine, 2009. Skin and non-skin probability approximation based on discriminative tree distribution. *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP'09)*, Piscataway, NJ, USA, pp: 2353-2356.
- Hassanpour, R., A. Shahbahrani and S. Wong, 2008. Adaptive Gaussian mixture model for skin color segmentation. *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 31.

- Jedynak, B., Z. Huicheng and D. Mohamed, 2003. Maximum entropy models for skin detection. *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer Berlin Heidelberg.
- Kakumanu, P., S. Makrogiannis and N. Bourbakis, 2007. A survey of skin-color modeling and detection methods. *Pattern Recogn.*, 40(3): 1106-1122.
- Kovac, J., P. Peer and F. Solina, 2003. Human skin color clustering for face detection. *Proceeding of EUROCON 2003*, 2: 144-148.
- Kruppa, H., A.B. Martin and S. Bernt, 2002. Skin patch detection in real-world images. *Ann. Sympos. Pattern Recogn.*, LNCS 2449:109-117.
- Ng, A.Y., M.I. Jordan and Y. Weiss, 2001. On spectral clustering: Analysis and an algorithm. In Dietterich, T., S. Becker and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*: 849-856.
- Phung, S., A. Bouzerdoum and D. Chai, 2005. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(1): 148-154.
- Pun, C. and P. Ng, 2014. A self-adaptive gaussian mixture model, 2014. *Comput. Vision Image Understanding*, 122(0): 35-46.
- Rehg, J.M. and M.J. Jones, 2002. Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1): 81-96.
- Skarbek, W., A. Koschan, T. Bercht, V. Zur and Klette, 1994. *Colour image segmentation-a survey*. CiteSeer 1994.
- Sun, H., 2010. Skin detection for single images using dynamic skin color modeling. *Pattern Recogn.*, 43(4):1413-1420.
- Yang, J., L. Weier and W. Alex, 1998. Skin-color modeling and adaptation. *Proceeding of 3rd Asian Conference on Computer Vision*, 2: 687-694.