

Skin Color Detection Model Using Neural Networks and its Performance Evaluation

¹K.K. Bhoyar and ²O.G. Kakde

¹Department of Computer and Information Technology,
Yeshwantrao Chavan College of Engineering, Nagpur, India

²Department of Electronics and Computer Science,
Vishweshwarayya National Institute of Technology, Nagpur, India

Abstract: Problem statement: Skin color detection is used as a preliminary step in numerous computer vision applications like face detection, nudity recognition, hand gesture detection and person identification. In this study we present a pixel based skin color classification approach, for detecting skin pixels and non skin pixels in color images, using a novel neural network symmetric classifier. The neural classifiers used in the literature either uses a symmetric model with single neuron in the output layer or uses two separate neural networks (asymmetric model) for each of the skin and non-skin classes. The novelty of our approach is that it has two output layer neurons; one each for skin and non-skin class, instead of using two separate classifiers. Thus by using a single neural network classifier we have improved the separability between these two classes, eliminating additional time complexity that is needed in asymmetric classifier. **Approach:** Skin samples from web images of people from different ethnic groups were collected and used for training. Ground truth skin segmented images were obtained by using semiautomatic skin segmentation tool developed by the authors. The ground truth database of skin segmented images, thus obtained was used to evaluate the performance of our NN based classifier. **Results:** With proper selection of optimum classification threshold that varies from image to image the classifier gave the detection rate of more than 90% with 7% false positives on an average, **Conclusion/Recommendations:** It is observed that the neural network is capable of detecting skin in complex lighting and background environments. The classifier has the ability to classify the skin pixels belonging to people from different ethnic groups even when they are present simultaneously in an image. The proper choice of optimum classification threshold that varies from image to image is an issue here. Automatic computation of this optimum threshold for each image is desired in practical skin detection applications. This issue can be taken up as a future study, which will enable us to perform fully automatic skin segmentation with reduced false positives.

Key words: Skin color, neural networks, RGB space, skin and non-skin pixels

INTRODUCTION

Skin color detection has been used in numerous computer vision applications like face detection, nudity recognition, hand gesture detection and person identification. Skin color detection is often used as a preliminary step in these applications. Color is the most robust and useful clue for skin detection and also allows fast processing of the skin patterns. Other cues like shape and geometry can be used to build accurate face detection systems.

Skin color detection is a challenging task as the skin color in an image is sensitive to various factors like illumination, camera characteristics, ethnicity,

individual characteristics such as age, sex and body parts and other factors like makeup, hairstyle and glasses. All these factors affect appearance of skin color (Kakumanu *et al.*, 2007). Another problem is that there is a significant overlap between the skin and non-skin pixels (Jones and Rehg, 2002).

Most of the skin detection techniques discussed in literature are used as a preprocessor for face detection and tracking systems. However when these techniques are used in real-time, it is crucial to follow time deadlines and memory constraints. Sometimes, accuracy may need to be sacrificed when the skin detection strategy is used only as a preprocessing step to face detection, particularly in real time applications.

Corresponding Author: K.K. Bhoyar, Department of Computer and Information Technology,
Yeshwantrao Chavan College of Engineering, Nagpur, India

In this study we have focused on the problem of developing an accurate and robust model for the human skin. The multilayer perception neural network with back propagation training algorithm has been used to build the classifier from a data set of skin pixels across different ethnic groups.

Review of skin color models using neural networks:

The Neural Network (NN) classifies the image regions as a collection of either skin or non-skin regions. Various approaches to skin modeling are used in the literature. Here we give a brief review of the neural network models for skin detection.

A Multi-Layer Perception (MLP) based skin color model for face detection is proposed by Ming-Jung *et al.* (2003). They have used 41000 skin pixels in RGB space having different illumination for training. Further they have used mask based processing to identify face regions, from the skin pixels identified. Karlekar and Desai (1999) and Phung *et al.* (2001) used MLP in CbCr space for skin classification. The MLP is trained from 200 images using Levenberg-Marquardt algorithm for faster convergence. Ming-Jung *et al.* (2003) and Zhu *et al.* (2004) trained a three layered NN in RGB space not only to extract the skin regions but also to interpolate the skin regions in 3D color cube. The NN interpolated area of the color cube is considered as skin region and the rest as non- skin region.

Two types of skin models are used in the literature (Brown *et al.*, 2001) viz., symmetric and asymmetric. Symmetric model uses a single classifier for both the classes whereas asymmetric model uses two separate classifiers for skin and non-skin pixels that are separately trained using respective features.

Brown *et al.* (2001) and Shin *et al.* (2002) trained two separate Self Organizing Maps (SOMs) from a set of about 500 manually labeled images, to learn skin-color and non-skin-color pixel distributions (asymmetric model).

Advantage of asymmetric skin classifier is that it increases the distances in certain skin related features between a positive (skin) and a negative (non-skin) image, with disadvantage of increased time complexity for training two classifiers. We have used symmetric approach using MLP neural network with back propagation training in RGB space.

The neural classifiers used in the literature either uses a symmetric model with single neuron in the output layer or uses two separate neural networks (asymmetric model) for each of the skin and non-skin classes. The novelty of our approach is that it has two output layer neurons; one each for skin and non-skin class, instead of using two separate classifiers. Thus by using a single NN

classifier we have improved the separability between these two classes, eliminating additional time complexity that is needed in asymmetric classifier.

MATERIALS AND METHODS

Neural network classifier:

Choice of a color space: A wide variety of color spaces with different properties are attributed to by developments in colorimetry, computer graphics and video signal transmission. The researchers in skin modeling have tried all these spaces in their study. The most popular among these spaces are, non-uniform color spaces like RGB, Normalized RGB (rbg), YCbCr, HSI, TSL and perceptually uniform color spaces like CIELAB and CIELUV. The choice of appropriate color space is often guided by the skin detection methodology and the application. It is to be noted here that the evaluation of color space goodness for skin modeling can not be performed because different modeling methods react very differently on the color space change (Vezhnevets *et al.*, 2003). It is well known that the illumination conditions of the scene clearly affect the color of the objects in the scene. The goal of any color-based system is to minimize this influence to make color-based recognition robust to illumination changes. It seems that chrominance-only color analysis should make the system somewhat independent from the lighting conditions. Hence many researchers have dropped the luminance component in order to take computational advantage. However, in (Shin *et al.* 2002) authors claimed that dropping the luminance component reduces the separability of skin and non-skin clusters. This is, of course, true because the projection of 3D data on a plane almost certainly smears skin and non-skin classes together. They further observed that the separability of skin and non-skin color classes is the highest in RGB space. Hence we have used RGB space for our experimentation.

Collection of training data: We have used collection of 500 web images comprising of different ethnic groups like Asian, African, Caucasian and Hispanic. 1000 samples each of skin and non-skin pixels are selected from this database for training. Few images containing people from different ethnic groups and skin types in the same image are deliberately chosen, so that the effectiveness of the classifier can be seen across all ethnic groups and skin types simultaneously.

The skin types typically include whitish, brownish, yellowish and darkish skins. Indoor and outdoor images with varying lighting conditions and varied backgrounds are also included in the database. The

representative skin colors are shown in Fig. 1 and the respective skin colors in RGB are shown in Table 1.

Neural network design: A key idea of NN is that after training, it is capable of generalizing from the training patterns and hence predicting the corresponding classes for patterns previously unknown to it. In other words the NN performs a high order of regression to fit the hidden function that relates its inputs (RGB pixel triplets here) and its outputs. One of the main advantages of using NN is that more complex partitioning of the feature space as shown in Fig. 3 is feasible by varying the network structure.

The architecture of the proposed three-layer feed forward neural network used for skin color classification is shown in Fig. 2. It is having three neurons in input layer, five neurons in hidden layer and two neurons in the output layer. The first neuron in the output layer represents skin class and the second neuron represents non-skin class.



Fig. 1: Typical skin colors from different ethnic groups and skin

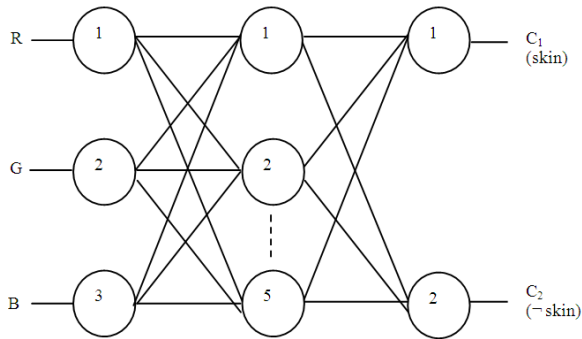


Fig. 2: Neural network architecture for skin/non-skin color classifier

Table 1: Respective skin colors in RGB for the colors in Fig. 1

(RGB)	(RGB)	(RGB)
(119,79,53)	(218,160,172)	(254,203,178)
(88,59,55)	(166,104,89)	(249,207,183)
(82,73,68)	(131,106,86)	(197,173,149)
(204,150,103)	(252,218,191)	(253,220,213)
(127,41,44)	(235,185,152)	(246,215,197)
(119,78,58)	(226,165,137)	(243,214,200)
(173,98,67)	(240,188,151)	(248,237,235)
(150,91,57)	(195,136,106)	(238,209,91)
(124,48,16)	(211,150,95)	(220,194,159)

A lot of experimentation has been done to find the number of neurons in the hidden layer of a 3×5×2 MLP network so as to achieve proper classification of the skin and non-skin samples. The hidden layer is required, as the patterns belonging to these two classes are linearly non-separable as seen in Fig. 3, which shows the skin color samples plotted in RGB cube as red points in 3-D space. It is to be noted that these are few representative skin pixels from the skin color space which is approximately 0.25% (Jones and Rehg, 2002) of the entire RGB space.

The network is trained using Error Back Propagation Training Algorithm (EBPTA). This algorithm minimizes the mean square error between the desired output and the actual output, using log sigmodal function for the hidden layer neurons and tan sigmodal function for the output layer neurons.

The neural network classifier has two outputs C_1 and C_2 , representing skin and non skin-pixel classes respectively. The output layer neuron has tan sigmoidal activation function. Hence C_1 and C_2 are in the interval (Vezhnevets *et al.*, 2003). Separate representative output neuron for each of these two classes gives better separability of skin and non-skin pixels. Ideally for skin pixels $C_1 - C_2 = 2$ and for non-skin pixels $C_1 - C_2 = -2$. To achieve a single step classification, a threshold θ ($0 < \theta \leq 2$) is introduced that can take care of pixels in the overlapping region containing skin and non-skin pixels. More value of $C_1 - C_2$ means, more confidence in the pixel as skin pixel. The classification here is done as follows:

$$\begin{aligned} \text{If } C_1 - C_2 > \theta & \text{ then pixel is skin pixel} \\ \text{else pixel is non - skin pixel} \end{aligned} \tag{1}$$

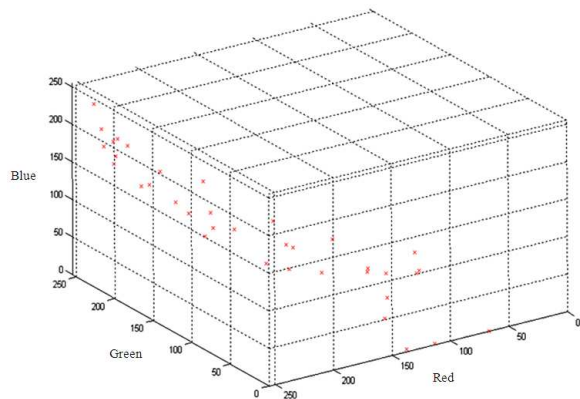
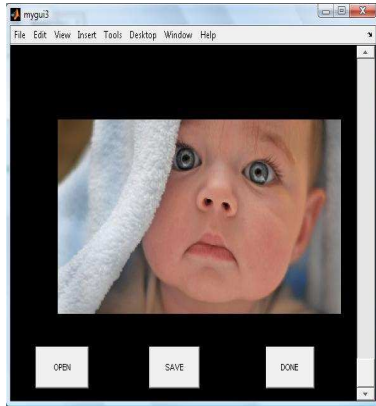
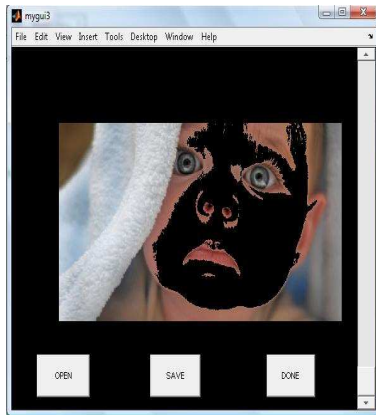


Fig. 3: Skin color samples plotted in red color in 3-D RGB space

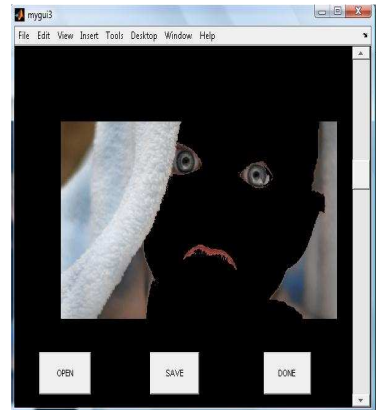
In Eq. 1, θ is a constant (threshold) for the given image, that plays a crucial role in controlling FPR and its optimum value is different for different images.



(a)



(b)



(c)

Fig. 4: Skin segmentation tool in action: Showing the screen immediately after loading image; (a): Partially segmented skin image; (b): And fully segmented image; (c): Identified skin pixels are shown in black here

The software tool developed in MATLAB 7 displays the image to be segmented in a window with GUI support as shown in Fig. 4. User clicks on a pixel within a skin area which is then taken as a seed pixel for 8-connected region growing algorithm. Threshold slider to the right may be adjusted to set appropriate.

Semi-automatic skin segmentation tool: This user interactive software tool developed by authors on the similar lines as that in (Jones and Rehg, 2002), can be used to segment skin pixels from an image. It is used to obtain perfectly skin segmented image interactively with the help of human expert. Such an image is called ground truth image and is used to evaluate the efficiency of other automatic segmentation algorithm in terms of False Positive Rate (FPR) and Detection Rate (DR), defined in the results and discussion.

RESULTS AND DISCUSSION

The results with $\theta = 0$ (default/auto detection) and with the best value of θ , for each image are presented here. All the skin pixels in the segmented images are marked with yellow color. For quantitative analysis of skin classifier we use False Positive Rate (FPR) and Detection Rate (DR). These are calculated using pixel by pixel comparison of perfectly segmented images (ground truth images) with their respective auto segmented images using Eq. 2 and 3:

$$FPR = \frac{\text{No. of non - skin pixels classified as skin}}{\text{Total no. of non - skin pixels}} \quad (2)$$

$$DR = \frac{\text{No. of skin pixels correctly classified}}{\text{Total no. of skin pixels}} \quad (3)$$

The drop in FPR for optimum value of threshold θ can be observed in Fig. 5-9 as the θ is changed from 0 to its optimum value. Figure 5 highlights one problem with color based skin model, that the colors which appears similar to skin colors, (here few whitish skin like pixels on the clothing) adds to FPR of the classifier.

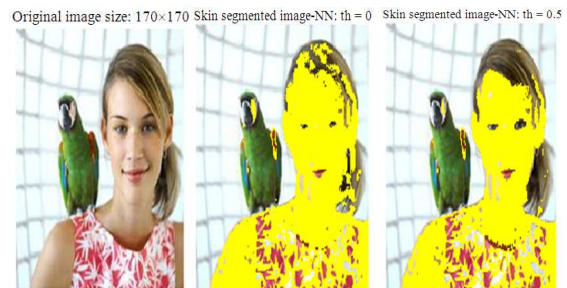


Fig. 5: (From left to right) original image-1, segmented image-1 with $\theta = 0$ and segmented image-1 with optimum threshold $\theta = 0.5$

Figure 6 shows the results for Image-2 which is noisy and contains two persons belonging to different ethnic groups. The skin segmented image here shows the success of this approach in spite of poor quality of original image.

Figure 7 shows the results for Image-3. Observe that the shadow portion on the neck has been correctly marked as skin region, although TPR is low in the region above the watch. Figure 8 gives the results for Image-4.

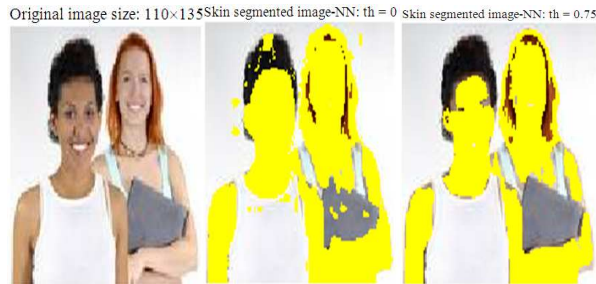


Fig. 6: (From left to right) original image-2, segmented image-2 with $\theta = 0$ and segmented image-2 with optimum threshold $\theta = 0.75$



Fig.7: (From left to right) original image-3, segmented image-3 with $\theta = 0$ and segmented image-3 with optimum threshold $\theta = 0.6$

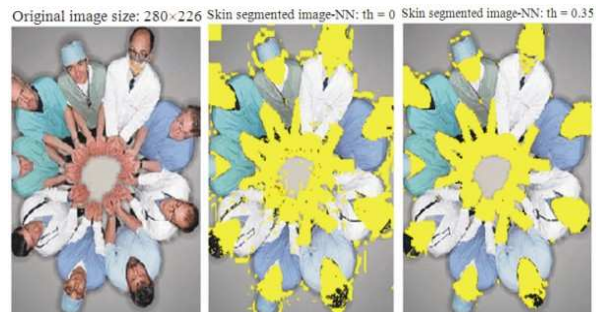


Fig. 8: (From left to right) original image-4, segmented image-4 with $\theta = 0$ and segmented image-4 with optimum threshold $\theta = 0.35$

Figure 9 shows almost perfect skin segmentation using optimum threshold. Similar results are obtained for other images in the database.

Figure 10 shows about 11.1% FPR in the first result and slightly reduced FPR (about 8.93%) with optimum threshold.

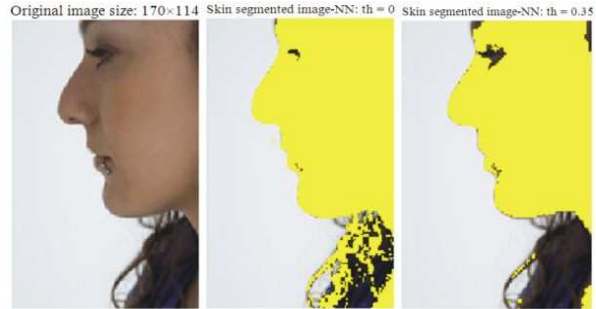


Fig. 9: (From left to right) original image-5, segmented image-5 with $\theta = 0$ and segmented image-5 with optimum threshold $\theta = 0.35$

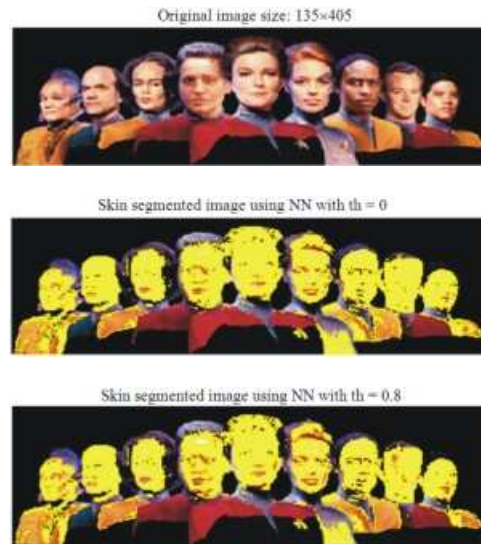


Fig. 10: (From top to bottom) original image-6, segmented image-6 with $\theta = 0$ and segmented image-6 with optimum threshold $\theta = 0.8$

Table 2: Qualitative analysis of NN skin classifier

Image	With $\theta = 0$		With optimum θ	
	DR (%)	FPR (%)	DR (%)	FPR (%)
Image-1	93.14	9.33	91.67	4.33
Image-2	96.10	9.66	95.31	3.00
Image-3	91.45	9.88	90.14	2.32
Image-4	98.01	7.31	99.49	3.63
Image-5	99.23	6.64	99.12	3.10
Image-6	88.00	11.10	86.00	8.93

CONCLUSION

It is observed that the classifier is invariant to illumination changes and works properly across the people from different ethnic groups. Although $\theta = 0$ gives acceptable results, it can be observed from the Table 2 that the False Positive Rate (FPR) is reduced by up to 7% when a particular θ (optimum value) is used for the given image. However the relation between image pixels and such optimum θ is not established in this study. Automatic computation of optimum threshold is desired in practical skin detection applications. Finding this relationship can be taken up as a future study, which will enable us to perform fully automatic skin segmentation with reduced FPR.

REFERENCES

- Brown, D.A., I. Craw and J. Lewthwaite, 2001. A SOM based approach to skin detection with application in real time systems. Proceeding of the British Machine Vision Conference, (BMVC'01), University of Leeds, UK., pp: 491-500. http://www.comp.leeds.ac.uk/bmvc2008/proceedings/2001/papers/33/accepted_33.pdf
- Kakumanu, P., S. Makrogiannis and N. Bporbakis, 2007. A survey of skin-color modeling and detection methods. *Patt. Recog.*, 40: 1106-1122. DOI: 10.1016/j.patcog.2006.06.010
- Karlekar, J. and U.B. Desai, 1999. Finding faces in color images using wavelet transform. Proceeding of the International Conference on Image Analysis and Processing, Sept. 27-29, IEEE Computer Society, Washington, DC., USA., pp: 1085-1088. DOI: 10.1109/ICIAP.1999.797744
- Jones, M.J. and J.M. Rehg, 2002. Statistical color models with applications to skin detection. *Int. J. Comput. Vis.*, 46: 81-96. DOI: 10.1023/A:1013200319198
- Ming-Jung, S., V. Deepthi and V.K. Asari, 2003. Neural network based skin color model for face detection. Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop, Oct. 15-17, IEEE Computer Society, USA., pp: 141-145. DOI: 10.1109/AIPR.2003.1284262
- Phung, S.L., D. Chai and A. Bouzerdoum, 2001. A universal and robust human skin color model using neural networks. Proceedings of the International Joint Conference on Neural Networks, July 2001, IEEE Xplore Press, Washington, DC., USA., pp: 2844-2849. DOI: 10.1109/IJCNN.2001.938827
- Shin, M.C., K.I. Chang and L.V. Tsap, 2002. Does color space transformation make any difference on skin detection. Proceedings of the 6th IEEE Workshop on Applications of Computer Vision, Dec. 3-4, IEEE Computer Society, Orlando, FL., pp. 275-279. DOI: 10.1109/ACV.2002.1182194
- Vezhnevets, V., V. Sazonov and A. Andreeva, 2003. A survey on pixel-based skin color detection techniques. Proceeding of the GraphiCon-2003, (GraphiCon'03), GraphiCon, UK., pp: 85-92. <http://graphicon.ru/oldgr/en/publications/text/gc2003vsa.pdf>
- Zhu, Q., C.T. Wu, K.T. Cheng and Y.L. Wu, 2004. An adaptive skin model and its applications to objectionable image filtering. Proceedings of the 12th Annual ACM International Conference on Multimedia, Oct. 10-16, ACM Press, New York, USA., pp: 56-63. DOI: 10.1145/1027527.1027538