Neural Network Based Skin Color Model for Face Detection

Ming-Jung Seow, Deepthi Valaparla, and Vijayan K. Asari
VLSI Systems Laboratory
Department of Electrical and Computer Engineering
Old Dominion University, Norfolk, VA 23529
Email: {mseow, dvala001, vasari}@odu.edu

Abstract

This paper presents a novel neural network based technique for face detection that eliminates limitations pertaining to the skin color variations among people. We propose to model the skin color in the three dimensional RGB space which is a color cube consisting of all the possible color combinations. Skin samples in images with varying lighting conditions, from the Old Dominion University skin database, are used for obtaining a skin color distribution. The primary color components of each plane of the color cube are fed to a three-layered network, trained using the backpropagation algorithm with the skin samples, to extract the skin regions from the planes and interpolate them so as to provide an optimum decision boundary and hence the positive skin samples for the skin classifier. The use of the color cube eliminates the difficulties of finding the non-skin part of training samples since the interpolated data is consider skin and rest of the color cube is consider non-skin. Subsequent face detection is aided by the color, geometry and motion information analyses of each frame in a video sequence. The performance of the new face detection technique has been tested with real-time data of size 320×240 frames from video sequences captured by a surveillance camera. It is observed that the network can differentiate skin and non-skin effectively while minimizing false detections to a large extent when compared with the existing techniques. In addition, it is seen that the network is capable of performing face detection in complex lighting and background environments.

1. Introduction

Unsupervised surveillance gadgets aided by hi-tech visual information indexing and retrieval systems are proving to be indispensable in the existing terror environment. Real-time applications such as frontal view face detection [1], tracking faces in natural scenes [2] and adult content filters [3] rely heavily on computationally inexpensive and adaptive face localization schemes which work in unconstrained environments. Face localization is the preliminary step towards recognition and identification of human facial characteristics in a visual scene. Contemporary approaches have built skin detectors based on histograms [4], which have the expensive time constraint. Skin color model based on the Gaussian distribution have not been able to deal with the variation of the skin distribution in non-uniform lighting environments [5]. Robust skin segmentation involves the formulation of an efficient mathematical model [6] to represent the skin color distribution. Segmenting skin from real-world images is a difficult task even though human skin is known to possess a unique color range, which is but a fraction of all the possible color combinations. The advantage of using the RGB color space in skin detection methods has been emphasized by Shin et al. [7]. in their comprehensive evaluation of existing color space transformations using four different separability criteria based on the scatter matrix and histogram analysis. Elimination of the illumination content by normalization [8] worsened the discriminating power due to projection of image data into lower dimensional space.

Face detection using statistical models in [9 and 10] have been strongly undermined by their inability to apply higher global constraints on the face template. In addition to that they have not been able to negate the influence of noise or changes in facial expressions on the extracted features. Neural network based face detection [11] tests for the existence of human faces using a retinally connected neural network, which requires scanning of the entire image for faces of different sizes and rotations thereby resulting in loss of valuable computation time. We propose a novel scheme in section 2 to model skin color, by training a three-layered network with skin and non-skin examples, using the back propagation algorithm. The trained network shown in Figure 1 is used to estimate the probable skin regions in a three-dimensional color cube and interpolate to provide a reasonable estimate of

Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop (AIPR'03)
0-7695-2029-4/03 $ 17.00 © 2003 IEEE

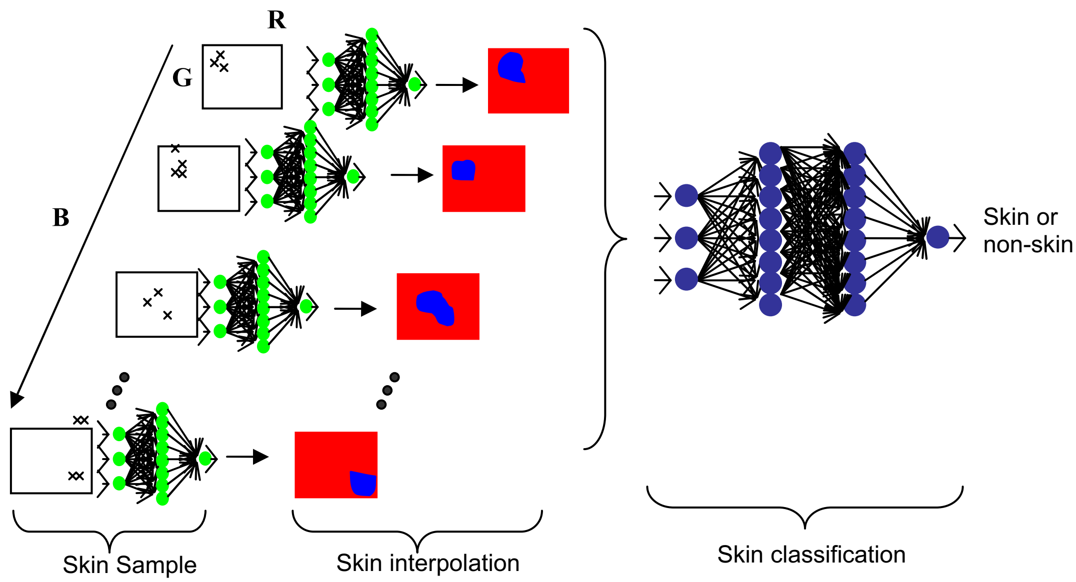


Figure1. Overview of the algorithm

the skin color distribution. A new technique for face detection has been proposed in section 3 that significantly enhances the computational speed for real-time applications. Section 4 provides a comprehensive look into the simulation and performance analysis of the proposed model. Section 5 concludes with an insight into the issues of human face tracking and improving the proposed model.

2. Modeling skin color in the RGB color space

Training a neural network for the skin color detection is challenging because of the difficulty in characterizing non-skin color. It is easy to get a representative sample of images, which contain skin color, but it is much harder to get a representative sample of those, which do not, since objects can have color similar to the skin color. The training of the skin color model based on neural networks composed of three stages: skin color collection, skin color interpolation, and skin color classification. Figure 1 illustrates the proposed method for training, extraction and classification using the skin color model.

2.1. Skin Database Compilation

Skin colors from various races of the world are collected from the Internet in the form of 10 × 10 pixels per skin sample for each individual from each image. 410 such samples were collected. As a result, there are 41,000 skin pixels having different illuminations in our skin color database used for training the neural network. The skin obtained was used to compute a three-dimensional histogram and assess the extent of skin color in the color cube comprising of all the possible color combinations. Figure 2 shows that the human skin is a fraction of the actual color cube, roughly about 0.25 % of the total colors present.

2.2. Skin color interpolation

Since the skin color samples we have collected do not represent the skin color population, we need to interpolate for the skin color that we don't have using the skin color we have.

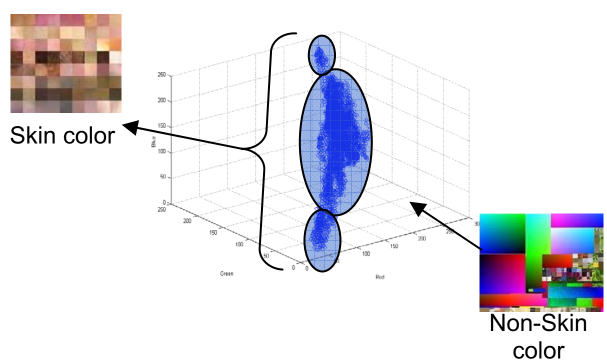


Figure 2: Histogram of the skin region

Multilayer perceptron trained using the back-propagation learning algorithm is used for skin color interpolation. Figure 3 illustrates the back propagation minimize the mean square error between the desired

Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop (AIPR'03)
0-7695-2029-4/03 $ 17.00 © 2003 IEEE

output and the actual output. The back propagation algorithm uses supervised learning where the training of the network is done with known input and output data. Once the network is trained, its weights can be used to compute outputs for new input values. We generated a 256×256×256 color cube to represent all the possible color combinations. Training the network involved learning of skin and non-skin examples. The primary color components of each of the 256 slices of the cube are fed to a three-layered network, trained using the back-propagation algorithm, to extract the skin regions (the crosses/ skin samples in Figure 1) from the slices and interpolate them so as to provide an optimized decision boundary for the skin regions (blue regions in Figure 1).

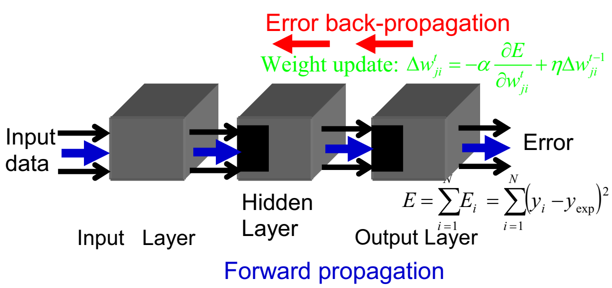


Figure 3: Back propagation algorithm

2.3. Skin color classification

Once the skin is interpolated it is used to train the three layer network for skin color classification. A three layered network is used since it is conceived from the color cube that the probable skin regions can be encapsulated by the complex decision boundaries such as the ones depicted in Figure 4, probably a hexagon.

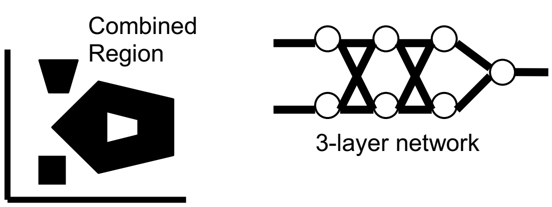


Figure 4: Optimized decision boundary

3. Face Detection

The proposed face detection technique analyzes the skin regions from each skin-segmented frame with a 20×20 mask similar to the one shown in Figure 5. Probable face regions in the images are divided into 9 regions according to the scheme shown in the figure. Nine different back propagation networks are used to decide the proximity of each of the 9 regions to a face region. The votes from each of the back propagation networks are used to decide if the probable face region is a face.

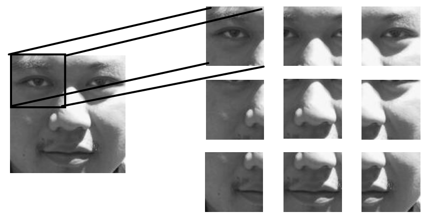


Figure 5. Architecture of the network

Result of the skin segmentation is used by the mask to scan for the face by the scanning process shown in Figure 6. Each of the 9 regions obtained by using the mask on a pixel-by-pixel scan of the image is subjected to a threshold function to decide if it corresponds to a facial feature. The presence of 4 facial features in the regions is needed for it to be detected as a face.

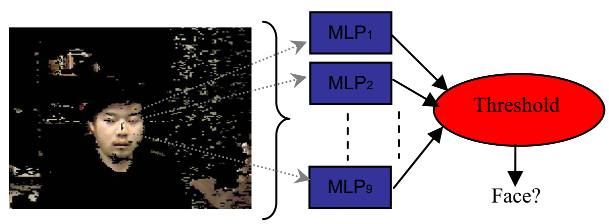


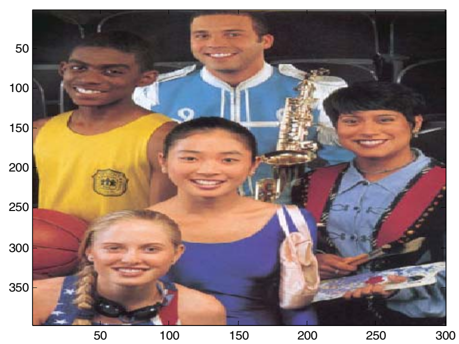
Figure 6. Illustration of the scanning process.

4. Simulation and Results

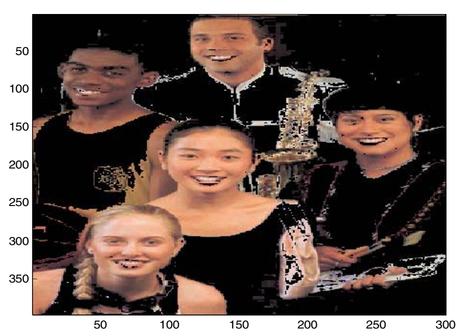
In addition to the skin samples that were collected during the skin database compilation, about 200 test images of individual persons and 150 images containing group photos were collected. Experiments were conducted in a Windows NT environment using Microsoft Visual C++ on a set of still images and video sequences. The first part of the simulation shows two still images with people of different races. The proposed method extracts skin very well irrespective of color of the skin as is evident from the results in Figures 7 (b) and 7 (d). The system was also used to detect and track human faces in real-time video sequences captured by an off the shelf Logitech camera, as is evident from Figure 8. It was able to detect faces in varying lighting environments too. The proposed face detection routine performed very well when compared to Rowley's method [11]. The reason being that the method does not need to scan the entire image every time. The skin detector reduces the search area considerably, saving a lot of computation time and enhancing the speed of the system.

7(a)

7(b)

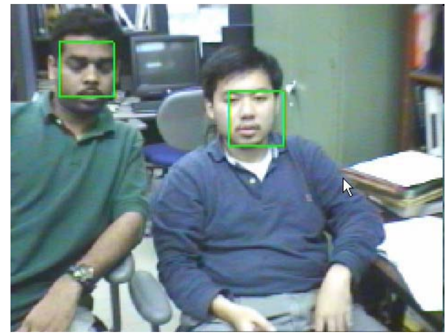


7(c)



7(d)

Figure 7. (a) Still test image 1, (b) Result (c) Image 2, (d) Result



8(a)

8(b)

8(c)

8(d)

Figure 8. (a - d) Result of face detection in a real-time video sequence

Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop (AIPR'03)
0-7695-2029-4/03 $ 17.00 © 2003 IEEE

5. Conclusion

A novel neural network based skin color model has been presented in this paper. The method has incorporated the concept of the color cube, back propagation and a novel filter based face detection technique to provide an accurate skin color model. The model has been used to segment still images and track faces in real-time video sequences in non-uniform lighting conditions. The computation time has been reduced considerably when compared to conventional techniques facilitating real-time applications. Further work is in progress to develop a face recognition system to identify and index individuals for surveillance purposes.

6. References

[1]. A. Eleftheriadis and A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates," *Signal Processing: Image Communication*, No. 7, 1995, pp. 231-248.

[2]. Y. Raja, S. J. McKenna and S. Gong, "Tracking and segmenting people in varying lighting conditions using color," *Proc. International Conference on Face and Gesture Recognition*, pp. 228-233, 1998.

[3]. M. M. Fleck, D. A. Forsyth and C. Dregler, "Finding naked people," *Proc. European Conference on Computer Vision*, 1996, pp. 593-602.

[4]. R. Kieldsen, J. Kender, "Finding skin color in images," *Proc. International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 312-317.

[5]. S.J McKenna, S. Gong, Y. Raja, Modelling facial colour and identity with Gaussian mixtures, *Pattern Recognition*, vol. 31, no. 12, 1998, pp. 1883–1892.

[6]. J.C. Terrillon, M. David, S. Akamatsu, "Automatic detection of human faces in natural scene images by use of a skin color model and invariant moments," *Proc. International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 112–117.

[7]. M. C. Shin, K. I. Chang, and L. V. Tsap, "Does Colorspace Transformation Make Any Difference on Skin Detection?," *IEEE Workshop on Applications of Computer Vision*, Orlando, FL, Dec 2002.

[8]. J. Yang and A. Waibel, "Tracking human faces in real-time," *Proc. IEEE Workshop on Applications of Computer Vision*, 1996.

[9]. K.C.Yow and R. Cipolla, "Feature-based human face detection," *Image and Vision Computing*, 1997, vol. 15, no. 9, pp. 713–735.

[10]. S. H. Jeng, Y. Y. M. Liao, C. C. Han, M. Y. Chern, Y. T. Liu, "Facial feature detection using geometrical face model: an efficient approach," *Pattern Recognition,* 1998, vol. 31, no. 3, pp. 273–282.

[11]. H. Rowley, S. Baluja, and T. Kanade, "Neural network based face detection," *IEEE Trans. Patt. Recog. Mach. Intell.*, January 1998, vol. 20, no. 1, pp. 23–38.