# Face Recognition Using Multispectral Random Field Texture Models, Color Content, and Biometric Features

Orlando J. Hernandez and Mitchell S. Kleiman
*Electrical and Computer Engineering*
*The College of New Jersey*
*Ewing, New Jersey 08628-0718*
*{hernande, kleiman2}@tcnj.edu*

## Abstract

*Most of the available research on face recognition has been performed using gray scale imagery. This paper presents a novel two-pass face recognition system that uses a Multispectral Random Field Texture Model, specifically the Multispectral Simultaneous Auto Regressive (MSAR) model, and illumination invariant color features. During the first pass, the system detects and segments a face from the background of a color image, and confirms the detection based on a statistically modeled skin pixel map and the elliptical nature of human faces. In the second pass, the face regions are located using the same image segmentation approach on a subspace of the original image, biometric information, and spatial relationships. The determined facial features are then assigned biometric values based on anthropometrics, and a set of vectors is created to determine similarity in the facial feature space.*

## 1. Introduction and background

With an increasing reliance on technology for biometric classification of individuals for security purposes, face recognition as a means of security, continues to be researched with new and varying approaches. While the majority of face recognition research has been investigated using grayscale images of human faces, there has been some previous work involving color images to detect or recognize the face [1]. The first known paper involving color detection of human faces used the chromaticity of individual pixels in an image to segment facial features, showing that chromaticity was consistent for skin, but that highlights and shadows in an image could pose problems using a chromaticity based approach [2].

The YUV, HSV, and RGB color spaces were investigated using variations of the Eigenface technique for the different color channels, but showed similar recognition results for all three color spaces [3]. Color and grayscale images were compared to determine if color held any advantage to grayscale, specifically using the Eigenface approach, and noted that while color held no significant advantage to grayscale using Eigenfaces, the red channel itself could improve performance because it was least sensitive to illumination changes [4]. Color images and grayscale images were also compared using luma-based maps for the individual facial features, creating fiducial points and using distances and angles between specified fiducial points to perform recognition. For the three chosen test sets from the XM2VTS database [5], the method showed between 90 and 95% recognition [6]. Neural networks based on skin color were also used, along with composite Principal Component Analysis (PCA), showing 90% recognition rate on a given test film [7]. Face morphing and light biasing techniques with Independent Component Analysis achieved 92% recognition [8]. Using a skin model based on 8 faces, segmenting through the YpbPr color space and a Support Vector Machine (SVM) for the eyes achieved 89.6% face detection in XM2VTS Database 1, and 75.9% detection in the XM2VTS Database 2 [9].

## 2. Multispectral random field texture model image segmentation

This paper primarily investigates the use of Multispectral Random Field Texture Models with regard to face recognition. A given image is analyzed using a sliding window that extracts 22 dimensional "Color Content Color Texture" ($C^3T$) features from

each window based on Multispectral Simultaneous Auto Regressive (MSAR) model and color features approach, clustering the windows using an unsupervised peak-climbing histogram algorithm [10]. The sliding window size is varied separately over a given range of values, and the best window size is automatically chosen based on the number of clusters resulting at each sliding window size. The resulting clusters are mapped back to the original image for analysis. Originally designed for rectangular images, the image segmentation algorithm was adapted to also perform segmentation using a mask for the image pixels.
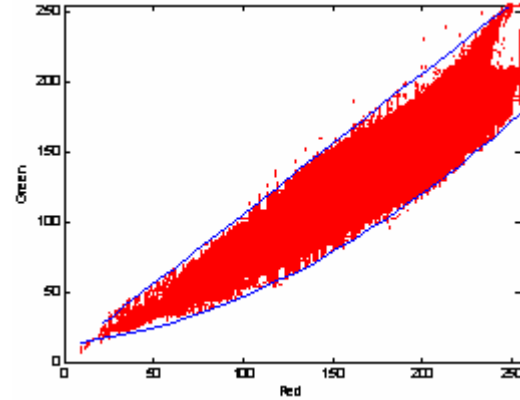
## 3. Skin map

Thirteen images from the FERET Color Database [11] and six from the XM2VTS Database were selected to create the skin map. The subjects were chosen so that the different type of possible skin colors would be contained in the sample set. The skin of the face was manually cropped and combined to a large image, which is shown in **Fig. 1**. In this image, all distinct pixel values (8-bit Red, Green, and Blue channels) were extracted and graphed.
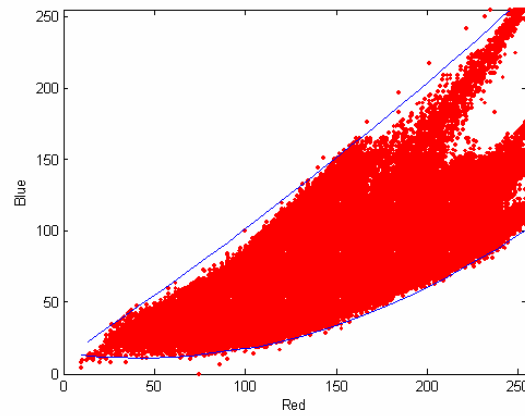


**Fig. 1.  Skin source**

The resulting skin pixel map was graphed using 2D views for GREEN vs. RED – **Fig. 2**(a), BLUE vs. RED – **Fig. 2**(b), and BLUE vs. GREEN – **Fig. 2**(c). The resulting upper and lower boundaries on each
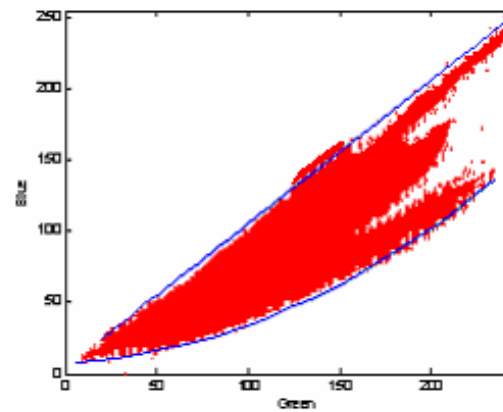
graph were modeled using best-fit linear, quadratic, or exponential functions. All pixels in a given image that were contained in between the upper and lower modeled equations in all three views were considered skin, otherwise the pixel was considered non-skin.



(a)



(b)



(c)

**Fig. 2.  Skin pixel mapping**

GREEN vs. RED:
$$\text{UPPER LIMIT} = \lceil 1.00102\,R + 5.76 \rceil \tag{1}$$

$$\text{LOWER LIMIT} = \left\lfloor \begin{array}{c} 0.002\,R^2 + 0.1366\,R \\ + 13.03 \end{array} \right\rfloor \tag{2}$$

BLUE vs. RED:
$$\text{UPPER LIMIT} = \left\lceil \begin{array}{c} 0.0006\,R^2 + 0.8445\,R \\ + 11.1554 \end{array} \right\rceil \tag{3}$$

$$\text{LOWER LIMIT} = \left\lfloor \begin{array}{c} 0.002\,R^2 + 0.1689\,R \\ + 14.584 \end{array} \right\rfloor \tag{4}$$

BLUE vs. GREEN:
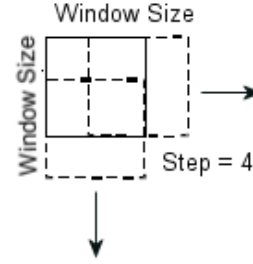$$\text{UPPER LIMIT} = \lceil 1.0058\,G + 4.6 \rceil \tag{5}$$

$$\text{LOWER LIMIT} = \left\lfloor \begin{array}{c} 0.0021\,G^2 \\ + 0.0483\,G \\ + 8.2725 \end{array} \right\rfloor \tag{6}$$

Based on the skin-map graphs, an anti-pixel was chosen to contain 0 Red, 255 Blue, and 255 Green (the color Cyan) due to the weighting of graphs towards Red, and away from Green and Blue. This anti-pixel is used to replace the pixels of designated non-skin areas.

## 4. Face separation

Images from the FERET and XM2VTS databases were initially segmented at all whole number ratios of the original image dimensions that were less than the original image size to determine the optimal image size to separate the face (FERET - 512 x 768 for a 2:3 ratio, XM2VTS - 720 x 576 for a 5:4 ratio). It was desired to have the resulting segmentation generate a solid region containing the face (defined as the eyes, nose, and mouth), and a small amount of additional regions containing the background, hair, or shirt. This criterion was used to determine the best possible size for both image databases, using a test set of 15 images. The sliding window of the image segmentation was also investigated, and the most effective choice were window sizes of 4, 8, and 12 pixels with a step size of 4 pixels, shown in **Fig. 3**. The lower of the window sizes that resulted in the closest amount of regions was the one chosen in the first-pass of segmentation. Window sizes larger than 12 pixels yielded a larger amount of different regions than desired, and a fixed window sizes (no range of window sizes) did not perform as effectively. The FERET database was resized to 122 x 183 pixels, and the XM2VTS database was resized to 190 x 152 pixels.



**Fig. 3. Clustering technique**

To improve segmentation results, the background of each image was removed by replacing square blocks of all non-skin pixels (based on the skin-map) with the anti-pixel before segmentation. Each database was tested using replacement block sizes from 1 to 20 pixels on 15 separate images from each database. It was shown that block sizes without overlap of 10 pixels for the FERET database and 6 pixels for the XM2VTS database were most effective to aiding image segmentation. Those blocks that had no adjacent blocks of completely non-skin were returned to their original pixel values. While the entire background could be successfully removed pixel by pixel in many cases with small block sizes, image segmentation was improved by jagged edges around the face and thus larger block sizes were chosen. The background removal is a tool to aid image segmentation; however, the system can function properly without removal and thus backgrounds with skin colors are capable of being segmented.

After background removal and image segmentation, the different cluster regions were fitted to best-fit ellipses determined by the three geometric moments from the centroid of each region. Ellipses were used to model the elliptical nature of human faces.

$$\mu_{xx} = \sum_{\text{Over Rows}} \sum_{\text{Over Columns}} \frac{(x - \bar{x})^2}{A} \tag{7}$$

$$\mu_{yy} = \sum_{\text{Over Rows}} \sum_{\text{Over Columns}} \frac{(y - \bar{y})^2}{A} \tag{8}$$

$$\mu_{xy} = \sum_{\text{Over Rows}} \sum_{\text{Over Columns}} \frac{(x - \bar{x})(y - \bar{y})}{A} \tag{9}$$

The ellipses were checked for the percentage non-skin pixels based on the skin-map, and the ellipse with the lowest non-skin percentage and above the minimum rectangular width of 68 pixels and rectangular height of 53 pixels was designated the extracted face from the image.

The corresponding region contained inside the face ellipse was converted to a pixel mask where a 0 indicated the given pixel was not contained in the region, and a 1 indicated that the given pixel was contained in the region. In any given row inside the ellipse, any 0 valued mask pixels that were contained in between 1 valued mask pixels were converted to 1s to create a more solid pixel mask. This pixel mask was resized to match the second pass image size and utilized in face recognition. A face was considered successfully separated if it contained fully the eyes, nose, and mouth, and segmented the entire image to more than a single region.

## 5. Face recognition

The original image size was utilized for the second pass of the image segmentation. Based on an enlarged pixel mask created from the first pass of the image segmentation, only those clusters of which the entire set of mask pixels were all set to 1 were considered for the second pass of the image segmentation.

Window sizes of 12, 16, 20, 24, and 28 pixels with step sizes of 4 were used for the second pass of the image segmentation for each image. Due to the high variability of the resulting image segmentation with the given test images, the derivative of the number of regions found at each of the window sizes was used to choose the best segmentation for the second pass of each face. Image segmentation for the peak and valley of the derivative graph were kept, and the rest discarded. If the peak or valley were 8 or 12, the larger window size of 12 was chosen because larger window sizes generally showed better results for feature separation.

After separation of the peak and valley images, the largest region was determined to be the skin of the face. Due to the large amount of smaller regions, all regions not corresponding to the skin of the face with pixels touching were merged to form larger possible facial feature regions. Those regions that had areas of less than 6 square clusters were merged with the main skin region and not considered as possible facial features. An example is shown in **Fig. 4**.



**Fig. 4. Example second-pass mapped back image**

The mouth was first identified in the peak and valley segmented images and used as a baseline for further identification and classification of facial features. The mouth was determined to appear between 25% and 45% of the height of the masked image. A mouth was detected if it fell within the given height range in the masked region, had an area in pixels greater than 12 window sizes squared, and had a size of less than 50% of the width of the masked image. If more than one region met the previous criteria, the region with a larger width was determined to be the mouth, and in the event a tie, the larger area region was chosen. The area of the mouth region in both the peak and valley images was calculated, and the image with the larger mouth area was utilized for recognition. In the event of a tie, the larger window size was used.

The centroid of the mouth for each image was calculated. A 5% span each direction vertical and 25% span each direction horizontal from the centroid was drawn, and the pixel values in the region were explored to find the dark pixels corresponding to the separation of the lips. The leftmost pixel in the span containing red components less than 140, and green and blue components less than 60 was used as the new leftmost point for the mouth region, and the rightmost pixel within the span meeting the dark pixel specifications was the rightmost point on the mouth. The entire process is shown in **Fig. 5**. If there are multiple pixels at the same x coordinates of the leftmost or rightmost points of the mouth, the y coordinate closest to the centroid is chosen. The centroid x coordinate was recalculated as the average of the x coordinates of these two points. If no mouth were found, the leftmost and rightmost dark pixel corresponding to the lips separation were found within the range between 25% and 45% of the height of the masked image, and the centroid was made equal to the midpoint of the two points.

**Fig. 5. Mouth determination**



**Fig. 6. Eye determination**

With the centroid of the mouth as a baseline, the segmented regions appearing at between 30% and 50% of the distance between the mouth and top of the masked image were investigated as possible eyes. Starting from the largest region to the smallest, each segmented region in the given area was searched with a span of 20% the masked image width on each side and 5% the masked image height on each side of the centroid for the presence of a dark circular region corresponding to the eye's pupil. The determined pupil contained only pixels falling below a value of 35 for the red, green, and blue components in the region. In the event that no pupil was detected in the span near the segmented regions in the eye area, the point at 25% the width of the face, and 40% the height between the mouth and top of the masked image was used to start the search for the left pupil. The dark pixels and spans are shown in **Fig. 6**. The centroid of the smallest detected pupil region with an area greater than 15 pixels was determined as the pupil centroid. If this pupil region was to the right of the x coordinate of the centroid of the mouth, it was considered to be the centroid of the left eye, and the right eye otherwise. Using the symmetry of the face, the point equidistant to the detected pupil region centroid was used as a starting point for the search for the other pupil using the same span distances.
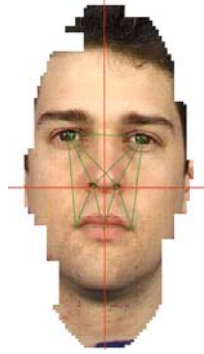
The nostrils of the nose were not usually separated directly in the image segmentation regions. Thus, based on the heuristics of the human face, the nose was shown to be a dark region near the average of the x and y centroids for the eyes and mouth. The nostrils were identified by small regions with every red component less than 140, and every green and blue component with values less than 60. Based on the x coordinate of the centroid of the mouth, the left nostril was identified as the centroid of the region meeting the criteria closest to the left of this midpoint of the face, and the right nostril identified by the centroid of the region meeting the criteria closest to the right midpoint of the face.

Anthropometrics were used with the centroids found for the left eye, right eye, left nostril, right nostril, and mouth, along with the leftmost point and rightmost point of the mouth for recognition. The midpoint of the nostrils was designated the coordinate (0, 0) of the face, and select anthropometric pixel distances between the points were used to characterize the face into a facial feature vector – **Fig. 7**. To test a face, the input image was put through the same two-pass segmentation process and the same vector of anthropometric distances was created. The similarity of this vector to each of the different vectors from the database was calculated based on the normalized Euclidian distance from each feature vector. Normalization was calculated by subtracting the value from the mean, and dividing by the sample variance, to yield a zero mean, unit variance distribution. The lowest difference within a given tolerance of 0.1 was considered a successful match; otherwise the face was not considered found in the database.

$$Normalized\ Distance = \frac{(x - \bar{x})}{\frac{1}{N-1}\sum(x - \bar{x})^2} \quad (10)$$

**Fig. 7. Selected anthropometric distances**

## 6. Results of face recognition

The similarity between the query image and a bank of images was tested using a small subset of the XM2VTS database and a leave-one-out approach. The face recognition system had 80% correct matches, 10% incorrect matches, and 10% non-matches exceeding the tolerances set forth for the Euclidian distance between vectors.

## 7. Conclusions

In this work, a color texture-based approach to face recognition was developed using a two-pass image segmentation approach based on Multispectral Random Field Texture Model. In the first pass of image segmentation, the face was first resized to a smaller size and put through a background removal process. Window sizes of 4, 8, and 12 pixels with a step size of 4 were used for the first pass, with the lower window size yielding the closest number of regions chosen as the first-pass segmentation results. The face of the resulting image was separated based on the lowest percentage non-skin pixels of the best-fit ellipses of the segmented regions. This region was used as a mask for the second pass of image segmentation. The mouth was first identified based on its location and size, and modified based on the middle portion between the lips. Then, the eyes were identified based on their position on the face and the darkness of the pupil. Lastly, the nostrils of the nose identified by its dark colored pixels and position in between the eyes and mouth. The centroids of the

nostrils, eyes, and mouth, along with the leftmost and rightmost points of the mouth were used to calculate anthropometric distances and create a feature vector for each face in the database. The similarity of the normalized feature vectors of a test image to images in the face database was used for recognition. The approach was demonstrated on a small image set.

## 8. References

[1] E. Hjelmas and B. K. Low, "Face Detection: A Survey", *Computer Vision and Image Understanding,* vol 83, 2001, pp. 236-274.

[2] T. C. Chang, T. S. Huang, and C. Novak, "Facial Feature Extraction from Color Images", *In Proc. The 12th IAPR International Conference on Pattern Recognition,* vol. 2, 1994, pp. 39-43.

[3] L. Torres, J. Y. Reutter, L. Lorente, "The Importance of Color Information in Face Recognition", *IEEE International Conference on Image Processing,* Kobe, Japan, October 25-28, 1999.

[4] S. Gutter, J. Hung, C. Liu, and H. Wechsler, "Comparative Performance Evaluation of Gray Scale and Color Information for Face Recognition Tasks", *3rd International Conference on Audio and Video Based Biometric Person Authentication,* AVBPA'01, Halmstad, Sweden, June 6-8, 2001.

[5] http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/

[6] R. Lanzarotti, "Facial Feature Detection and Description", *Ph.D. Thesis,* LAIV Laboratory, Università degli Studi di Milano, 2003.

[7] M.-J. Seow, R. Gottumukkal, D. Valaparla, and K. V. Asari, "A Robust Face Recognition System for Real Time Surveillance," *Proceedings of the IEEE Computer Society International Conference on Information Technology: Coding and Computing – ITCC 2004,* Las Vegas, Nevada, April 5-7, 2004, vol. 1, pp. 631-636.

[8] X. Yu and G. Baciu, "Face Recognition from Color Images in Presence of Dynamic Orientations and Illumination Conditions", ICBA 2004, pp. 227-233.

[9] P. Campadelli, R. Lanzarotti, and G. Lipori, "Face Detection in Color Images of Generic Scenes", *In Proceeding of the International Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2004).* pp. 97-103.

[10] A. Khotanzad and O. J. Hernandez, "Color Image Retrieval Using Multispectral Random Field Texture Model & Color Content Features", *Pattern Recognition Journal,* Volume 36, Issue 8, August 2003, pp. 1679 – 1694.

[11] http://www.nist.gov/humanid/colorferet