Skin Segmentation Using Color Pixel Classification: Analysis and Comparison

Son Lam Phung, *Member*, *IEEE*,
Abdesselam Bouzerdoum, *Sr. Member*, *IEEE*,
and Douglas Chai, *Sr. Member*, *IEEE*

Abstract—This paper presents a study of three important issues of the color pixel classification approach to skin segmentation: color representation, color quantization, and classification algorithm. Our analysis of several representative color spaces using the Bayesian classifier with the histogram technique shows that skin segmentation based on color pixel classification is largely unaffected by the choice of the color space. However, segmentation performance degrades when only chrominance channels are used in classification. Furthermore, we find that color quantization can be as low as 64 bins per channel, although higher histogram sizes give better segmentation performance. The Bayesian classifier with the histogram technique and the multilayer perceptron classifier are found to perform better compared to other tested classifiers, including three piecewise linear classifiers, three unimodal Gaussian classifiers, and a Gaussian mixture classifier.

Index Terms—Pixel classification, skin segmentation, classifier design and evaluation, color space, face detection.

1 Introduction

In recent years there has been a growing interest in the problem of skin segmentation, which aims to detect human skin regions in an image. Skin segmentation is commonly used in algorithms for face detection [1], [2], hand gesture analysis [3], and objectionable image filtering [4]. In these applications, the search space for objects of interest, such as faces or hands, can be reduced through the detection of skin regions. To this end, skin segmentation is very effective because it usually involves a small amount of computation and can be done regardless of pose.

Most existing skin segmentation techniques involve the classification of individual image pixels into skin and nonskin categories on the basis of pixel color. The rationale behind this approach is that the human skin has very consistent colors which are distinct from the colors of many other objects. In the past few years, a number of comparative studies of skin color pixel classification have been reported. Jones and Rehg [4] created the first large skin database—the Compaq database—and used the Bayesian classifier with the histogram technique for skin detection. Brand and Mason [5] compared three different techniques on the Compaq database: thresholding the red/green ratio, color space mapping with 1D indicator, and RGB skin probability map. Terrillon et al. [6] compared Gaussian and Gaussian mixture models across nine chrominance spaces on a set of 110 images of 30 Asian and Caucasian people. Shin *et al.* [7] compared skin segmentation in eight color spaces. In their study, skin samples were taken from the AR and the University of Oulo face databases and nonskin samples were taken from the University of Washington image database.

- *S.L. Phung and D. Chai are with the School of Engineering and Mathematics, Edith Cowan University, WA 6027, Australia. E-mail: {s.phung, douglas.chai}@ieee.org.*
- *A. Bouzerdoum is with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, NSW 2522, Australia. E-mail: a.bouzerdoum@ieee.org.*

Manuscript received 25 Mar. 2003; revised 20 May 2004; accepted 3 June 2004. Recommended for acceptance by Dr. Tan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0004-0303.

In this paper, we present a comprehensive study of three important issues of the color pixel classification approach to skin segmentation, namely color representation, color quantization, and classification algorithm. We investigate eight different color representations, seven different levels of color quantization, and nine different color pixel classification algorithms. To support this study, we have created a large image database consisting of 4,000 color images together with manually prepared ground-truth for skin segmentation and face detection. The paper is organized as follows: color representations and color pixel classification algorithms are described in Section 2, results of our analysis and comparison are presented in Section 3, and conclusions are given in Section 4.

2 Skin Color Classification

The aim of skin color pixel classification is to determine if a color pixel is a skin color or nonskin color. Good skin color pixel classification should provide coverage of all different skin types (blackish, yellowish, brownish, whitish, etc.) and cater for as many different lighting conditions as possible. This section describes the color spaces and the classification algorithms that will be investigated in this study.

2.1 Color Representations

In the past, different color spaces have been used in skin segmentation. In some cases, color classification is done using only pixel chrominance because it is expected that skin segmentation may become more robust to lighting variations if pixel luminance is discarded. In this paper, we investigate how the choice of color space and the use of chrominance channels affect skin segmentation. We should note that there exist numerous color spaces but many of them share similar characteristics. Hence, in this study, we focus on four representative color spaces which are commonly used in the image processing field [8]:

- **RGB**: Colors are specified in terms of the three primary colors: red (R), green (G), and blue (B).
- **HSV**: Colors are specified in terms of hue (H), saturation (S), and intensity value (V) which are the three attributes that are perceived about color. The transformation between HSV and RGB is nonlinear. Other similar color spaces are HIS, HLS, and HCI.
- **YCbCr**: Colors are specified in terms of luminance (the Y channel) and chrominance (Cb and Cr channels). The transformation between YCbCr and RGB is linear. Other similar color spaces include YIQ and YUV.
- **CIE-Lab**: Designed to approximate perceptually uniform color spaces (UCSs), the CIE-Lab color space is related to the RGB color space through a highly nonlinear transformation. Examples of similar color spaces are CIE-Luv and Farnsworth UCS.

2.2 Classification Algorithms

Several algorithms have been proposed for skin color pixel classification. They include piecewise linear classifiers [9], [10], [11], the Bayesian classifier with the histogram technique [4], [12], Gaussian classifiers [2], [13], [14], [15], and the multilayer perceptron [16]. The decision boundaries of these classifiers range from simple shapes (e.g., rectangle and ellipse) to more complex parametric and nonparametric forms.

2.2.1 Piecewise Linear Decision Boundary Classifiers

In this category of classifiers, skin and nonskin colors are separated using a piecewise linear decision boundary. For example, Chai and Ngan [9] proposed a face segmentation algorithm for a videophone application in which a fixed-range skin color map in the CbCr

TABLE 1
Data Sets of the ECU Database

Dataset	Description	Images
1	Original color images	4,000
2	Face segmented images	4,000
3	Skin segmented images	4,000
4	Cropped 64×64 face patterns	12,000
5	Nonface landscape images	2,000

plane is used. Sobottka and Pitas [11] proposed a set of fixed skin thresholds in the HS plane. These two approaches are based on the observation that skin chrominance, even across different skin types, has a small range, whereas skin luminance varies widely. Garcia and Tziritas [10] constructed a more complex skin color decision boundary that is made up of eight planes in the YCbCr space.

2.2.2 Bayesian Classifier with the Histogram Technique

The Bayesian decision rule for minimum cost is a well-established technique in statistical pattern classification [17]. Using this decision rule, a color pixel $\mathbf{x}$ is considered as a skin pixel if

$$\frac{p(\mathbf{x}|skin)}{p(\mathbf{x}|nonskin)} \geq \tau, \tag{1}$$

where $p(\mathbf{x}|skin)$ and $p(\mathbf{x}|nonskin)$ are the respective class-conditional pdfs of skin and nonskin colors and τ is a threshold. The theoretical value of τ that minimizes the total classification cost depends on the a priori probabilities of skin and nonskin and various classification costs; however, in practice τ is often determined empirically. The class-conditional pdfs can be estimated using histogram or parametric density estimation techniques. The Bayesian classifier with the histogram technique has been used for skin detection by Wang and Chang [12] and Jones and Rehg [4].

2.2.3 Gaussian Classifiers

The class-conditional pdf of skin colors is approximated by a parametric functional form, which is usually chosen to be a unimodal Gaussian [2], [14] or a mixture of Gaussians [13], [15]. In the case of the unimodal Gaussian model, the skin class-conditional pdf has the form:

$$\begin{aligned} p(\mathbf{x}|skin) &= g(\mathbf{x}; \mathbf{m}_s, \mathbf{C}_s) \\ &= (2\pi)^{-d/2} |\mathbf{C}_s|^{-1/2} exp\left\{ -\frac{1}{2}(\mathbf{x}-\mathbf{m}_s)^T \mathbf{C}_s^{-1} (\mathbf{x}-\mathbf{m}_s) \right\}, \end{aligned} \tag{2}$$

TABLE 2
Statistics of the Skin Data Set

Skin Types	Images
whitish, pinkish	1,665
yellowish, light brownish	1,402
reddish, darkish, dark brownish	965
other skin types	102
Lighting Conditions	**Images**
indoor lighting conditions	1,931
outdoor lighting conditions	1,855
other lighting conditions	214

where d is the dimension of the feature vector, $\mathbf{m}_s$ is the mean vector and $\mathbf{C}_s$ is the covariance matrix of the skin class. If we assume that the nonskin class is uniformly distributed, the Bayesian rule in (1) reduces to the following: a color pixel $\mathbf{x}$ is considered as a skin pixel if

$$(\mathbf{x}-\mathbf{m}_s)^T \mathbf{C}_s^{-1} (\mathbf{x}-\mathbf{m}_s) \leq \tau, \tag{3}$$

where τ is a threshold and the left hand side is the squared Mahalanobis distance. The resulting decision boundary is an ellipse in 2D space and an ellipsoid in 3D space. In this study, we also investigate the approach of modeling both skin and nonskin distributions as unimodal Gaussians. In this case, it can easily be shown that $\mathbf{x}$ is a skin pixel if

$$(\mathbf{x}-\mathbf{m}_s)^T \mathbf{C}_s^{-1} (\mathbf{x}-\mathbf{m}_s) - (\mathbf{x}-\mathbf{m}_{ns})^T \mathbf{C}_{ns}^{-1} (\mathbf{x}-\mathbf{m}_{ns}) \leq \tau, \tag{4}$$

where τ is a threshold and $\mathbf{m}_{ns}$ and $\mathbf{C}_{ns}$ are the mean and the covariance of the nonskin class, respectively. Another approach is to model both skin and nonskin distributions as Gaussian mixtures [13], [15]:

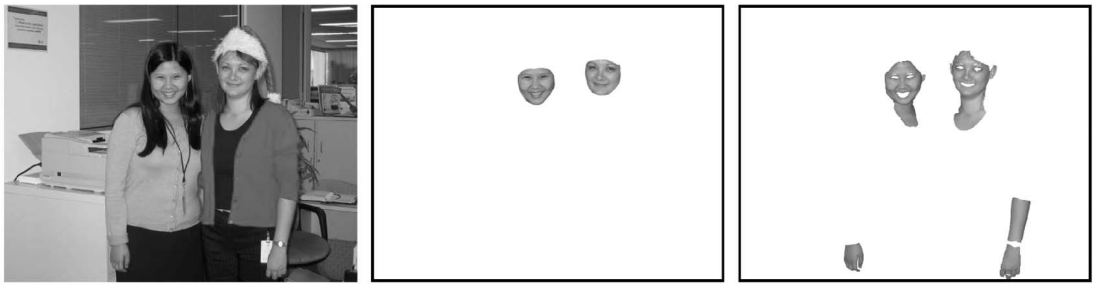
$$p(\mathbf{x}|skin) = \sum_{i=1}^{N_s} \omega_{s,i}\, g(\mathbf{x}; \mathbf{m}_{s,i}, \mathbf{C}_{s,i}), \tag{5}$$

$$p(\mathbf{x}|nonskin) = \sum_{i=1}^{N_{ns}} \omega_{ns,i}\, g(\mathbf{x}; \mathbf{m}_{ns,i}, \mathbf{C}_{ns,i}), \tag{6}$$

The parameters of a Gaussian mixture (i.e., weights ω, means $\mathbf{m}$, covariances $\mathbf{C}$) are typically found using the Expectation/Maximization algorithm.

2.2.4 Multilayer Perceptrons

The multilayer perceptron (MLP) is a feed-forward neural network that has been used extensively in classification and regression. A comprehensive introduction to the MLP can be found in [18].



(a) (b) (c)

Fig. 1. Sample images from the ECU face and skin detection database. (a) Original image (Set 1). (b) Face segmented image (Set 2). (c) Skin segmented image (Set 3).

TABLE 3
Skin Segmentation Data Sets for Training and Testing

Datasets	No. Images	Skin Pixels	Nonskin Pixels
Training	2,500	116.6 million	564.1 million
Testing	1,500	92.8 million	337.7 million

Compared to the piecewise linear or the unimodal Gaussian classifiers, the MLP is capable of producing more complex decision boundaries. In [16], we used the MLP to classify skin and nonskin pixels in the CbCr plane. In that work, only 100 images were used for training and testing the MLP. In this paper, the MLP technique is extended to a 3D color space and tested on a much larger data set.

3 Experimental Results and Analysis

A comprehensive set of experiments was performed to analyze the effects of color representation and color quantization on skin segmentation and to compare different classification algorithms. Before delving into the analysis and comparison parts, we first explain the data preparation process.

3.1 ECU Face and Skin Detection Database

The data was taken from the ECU face and skin detection database that we created at Edith Cowan University. The database has five data sets (see Table 1). Set 1 consists of 4,000 original color images: about 1 percent of these images were taken with our digital cameras and the rest were collected manually from the Web over a period of 12 months in 2002-2003. The image sources are too numerous to list here but they were chosen to ensure the diversity in terms of the background scenes, lighting conditions, and face and skin types. The lighting conditions include indoor lighting and outdoor lighting; the skin types include whitish, brownish, yellowish, and darkish skins. Sets 2 and 3 contain the respective face and skin detection ground-truth for the images in Set 1. The ground-truth images were meticulously prepared by manually segmenting the face and skin regions. The skin segmented images consist of all exposed skin regions such as facial skin, neck, arms, and hands. A rough categorization of the data set into different groups of skin types and lighting conditions is given in Table 2. Set 4 consists of 12,000 frontal-upright face patterns that we manually cropped from Web images, whereas Set 5 consists of 2,000 large landscape photos. Sample images from the database are shown in Fig. 1.

3.2 Analysis of Skin Color Pixel Classifiers

The data used for training and testing in our experiments are summarized in Table 3. For training, skin pixels were taken from skin segmented images and nonskin pixels from the complements of the skin segmented images. For testing, the skin color pixel classifiers were applied to the test images; no extra postprocessing was used. Each output image generated by a classifier was compared pixel wise with the corresponding skin segmented ground-truth. The segmentation performance was measured in terms of the correct detection rate *(CDR)*, the false detection rate *(FDR)*, and the overall classification rate *(CR)*. The CDR is the percentage of skin pixels correctly classified; the FDR is the percentage of nonskin pixels incorrectly classified; the CR is the percentage of pixels correctly classified.

In our study, nine skin color pixel classifiers were compared. These classifiers are summarized in Table 4. For the three piecewise linear classifiers, we took the fixed parameters directly from the original references [9], [10], [11]. For the Gaussian mixture classifier, we used the model parameters published by Jones and Rehg [4]. For the other five classifiers whose parameters were not available to us, we constructed the classifiers using our training data. Three unimodal Gaussian classifiers were tested: a 2D Gaussian classifier of skin in the CbCr plane, a 3D Gaussian classifier of skin in the YCbCr space, and a classifier with 3D Gaussians for both skin and nonskin the YCbCr color space. For the Bayesian classifier with the histogram technique, we used the RGB color space and histograms with 256^3 bins. The three unimodal Gaussian classifiers and the Bayesian classifier were constructed using the entire training set of 680.7 million samples presented in Table 3. For the MLP classifier, we extracted a training set of 30,000 skin and nonskin samples and trained the network using the Levenberg-Marquardt algorithm. Different network sizes and activation functions were investigated but we only report the performance of the best network.

The ROC curves and the classification rates of the tested classifiers are shown in Fig. 2 and Table 5, respectively. The Bayesian and MLP classifiers were found to have very similar performance. The Bayesian classifier had a maximum CR of 89.79 percent, whereas the MLP classifier had a maximum CR of 89.49 percent. Both classifiers performed consistently better than the Gaussian classifiers and the piecewise linear classifiers. Among the four Gaussian classifiers, the unimodal Gaussian classifier of both skin and nonskin (*3DG pos/neg*) had the best performance. This result shows that classification performance is improved if nonskin samples are also used in training the Gaussian models. Furthermore, the 3D unimodal Gaussian of skin (*3DG-pos*) outperformed its 2D counterpart (*2DG-pos*). The comparative performances of 3D and 2D feature vectors will be further examined in

TABLE 4
Tested Skin Color Pixel Classifiers

ID	Classifier	Classifier Parameters	Color Representation
CbCr-fixed	CbCr fixed-range: $77 \leq \text{Cb} \leq 127$ and $133 \leq \text{Cr} \leq 173$	[9]	CbCr
HS-fixed	HS fixed-range: $0.23 \leq \text{S} \leq 0.68$ and $0 \leq \text{H} \leq 50^o$	[11]	HS
GT plane-set	Garcia & Tziritas' plane set: skin cluster by 8 planes in YCbCr	[10]	YCbCr
Bayesian	Bayesian classifier with the histogram technique: 256^3 bins	trained	RGB
2DG-pos	2-D unimodal Gaussian of skin	trained	CbCr
3DG-pos	3-D unimodal Gaussian of skin	trained	YCbCr
3DG-pos/neg	3-D unimodal Gaussians of skin and nonskin	trained	YCbCr
3DGM	3-D Gaussian mixture of skin and nonskin	[4]	RGB
MLP	Multilayer perceptron	trained	RGB

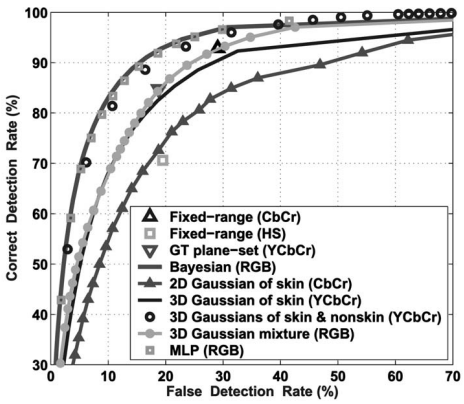


Fig. 2. ROC curves of skin color pixel classifiers.

Section 3.3. We found that the 3D Gaussian mixture classifier (*3DGM*) did not perform as well as the 3D unimodal Gaussian classifier of skin and nonskin. However, we should reiterate that the Gaussian mixture classifier was not trained with our training set; its parameters were taken from [4].

The thresholds of the three piecewise linear classifiers were fixed; hence, the corresponding ROC curves had only one point. These classifiers have an advantage in that they are all simple and fast. However, their performances were not as good as the Bayesian classifier, the MLP classifier, or the 3D skin and nonskin unimodal Gaussian classifier. The respective classification rates of the CbCr fixed-range classifier, the HS fixed-range classifier, and the GT plane-set classifier were 75.64 percent, 78.38 percent, and 82.00 percent.

In terms of memory usage, the Bayesian classifier using the histogram technique required the largest amount of memory. For example, each histogram in the Bayesian classifier used in this experiment has about 16.8 million entries. In comparison, the MLP classifiers that we trained have between nine and 15 neurons and fewer than 40 connections. Therefore, the MLP classifier is a good candidate if low memory usage is also a requirement. The 2D unimodal Gaussian of skin is characterized by 6 scalar parameters, the 3D unimodal Gaussian of skin by 12 parameters, and the 3D unimodal Gaussian model of skin and nonskin by 24 parameters. The Gaussian mixture model used in our test has 112 parameters. Finally, each of the two fixed-range classifiers has two parameters, whereas the GT plane-set classifier has 24 parameters.

Fig. 3 shows sample outputs of the Bayesian classifier with the histogram technique, the 3D unimodal Gaussian classifier with both skin and nonskin pdfs and the multilayer perceptron. The Bayesian and the MLP classifiers have almost similar segmentation outputs; they both make fewer false detections compared to the Gaussian classifier. The figure shows that Bayesian and MLP classifiers can successfully identify exposed skin regions including face, hands, and neck. However, objects in the background with similar colors as the skin will invariably lead to false detections, hence the need for postprocessing steps. Garcia and Tziritas [10] proposed a region growing technique whereby adjacent and similar skin-colored regions are merged; Fleck et al. [19] considered only skin-colored pixels that have small texture amplitude as skin pixels. However, a detailed treatment of postprocessing techniques is beyond the scope of this paper.

3.3 Analysis of Color Representations

We used the Bayesian classifier with the histogram technique to analyze different color representations. There are several reasons for this decision. Most importantly, using the histogram technique for pdf estimation, we do not need to make any assumption about the form of skin and nonskin densities. In contrast, if a particular form of the class-conditional pdf is assumed, as with the Gaussian density models, some color spaces may be favored over others. Furthermore, in this particular problem the feature vector has a low dimension and a large data set is available. Therefore, it is feasible to use the histogram technique for pdf estimation. Lastly, the Bayesian classifier with the histogram technique can be constructed very rapidly, even with a large training set, compared to other classifiers such as the MLP.

We analyzed a total of eight different feature vectors: RGB, HSV, YCbCr, CIE-Lab, normalized rg, HS, CbCr, and ab. The first four vectors consist of all color channels; the other four vectors consist of only chrominance channels. The analysis was carried out in seven different dyadic histogram sizes: 4, 8, 16, 32, 64, 128, and 256.

The ROC curves of the 8 feature vectors, with histogram sizes of 256 bins and 64 bins per channel, are shown in Fig. 4; the classification rates (CRs) at selected points on the ROC curves are shown in Table 6. We observe that, at the histogram size of 256 bins per channel, the classification performance was almost the same for the four color spaces tested, RGB, HSV, YCbCr, and CIE-Lab. This observation also holds for histogram sizes of 128 and 64 bins per channel. Therefore, we conclude that skin color pixel classification can be done in any of these color spaces. The most appropriate color space to use should depend on the input image format and the need of subsequent image processing steps. This result is expected because theoretically the overlap between skin and nonskin colors should not be affected by any one-to-one color space transformation. It is likely that the performance difference between color spaces is an effect of color quantization (i.e., histogram size). At high histogram sizes, the difference is very minor. Our conclusion agrees with that of Shin et al. [7], who measured the skin and nonskin separability in different color spaces using metrics derived from the class scatter matrices and histograms.

Results in Fig. 4 and Table 6 show that feature vectors comprising all channels (i.e., RGB, HSV, YCbCr, and CIE-Lab) outperformed feature vectors containing only chrominance

TABLE 5
Classification Rates (CRs) of Skin Color Pixel Classifiers

Classifier ID	CbCr-fixed	HS-fixed	GT plane-set	Bayesian	2DG-pos	3DG-pos	3DG-pos/neg	3DGM	MLP
FDR=10%	FDR=29.09	FDR=19.48	FDR = 18.77	88.75	82.37	85.27	88.01	85.23	88.46
FDR=15%	CR=75.64	CR=78.38	CR=82.00	86.17	81.07	83.45	85.57	83.63	85.97
FDR=20%	*fixed*	*fixed*	*fixed*	82.97	78.85	80.84	82.47	81.29	82.84
CR_{max}				89.79	82.67	85.57	88.92	85.76	89.49
99% conf. int. of CR_{max}	[75.58, 75.70]	[78.32 78.44]	[81.94 82.06]	[89.74 89.84]	[82.61 82.73	[85.52 85.62]	[88.87 88.97]	[85.71 85.81]	[89.44 89.54]

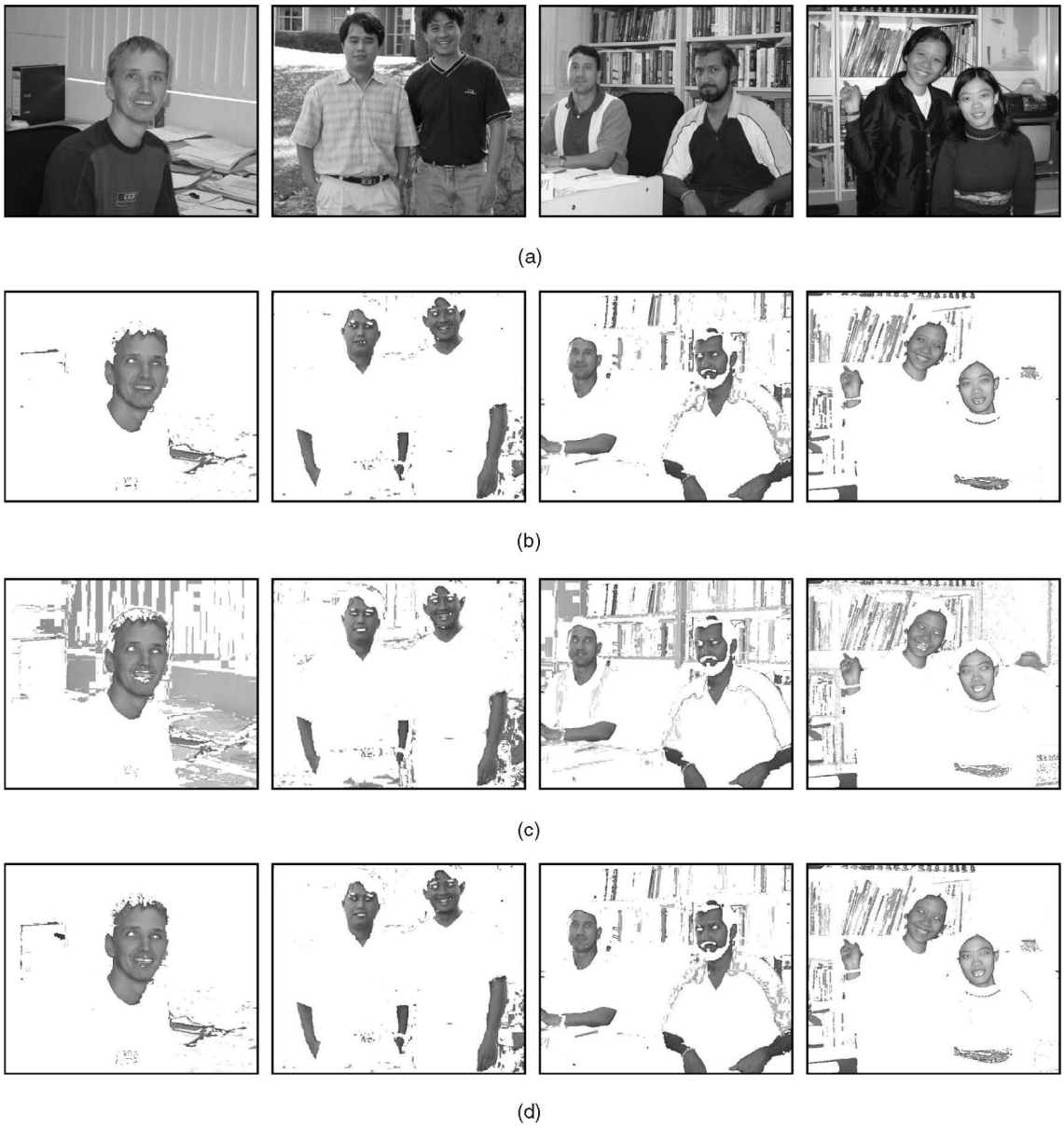


Fig. 3. Sample results of skin segmentation using color pixel classification. The classifier thresholds are chosen at the ROC curve point where FDR = 15 percent. (a) Input images. (b) Bayesian classifier with the histogram technique (64 bins per channel). (c) 3D unimodal Gaussian classifier (both skin and nonskin pdfs). (d) Multilayer perceptron.

channels (i.e., normalized rg, HS, CbCr, and ab). Computed over the four 3D feature vectors, the maximum CR had a mean of 89.61 percent and a standard deviation (*std*) of 0.22 percent; computed over the four 2D feature vectors, the maximum CR had a mean of 85.45 percent and a std of 0.97 percent. It has been previously suggested that skin detection can be made more robust to the lighting intensity if pixel luminance is not used in classification. However, such robustness to lighting intensity is essentially the result of expanding the skin color decision boundary to cover the entire luminance channel. As our results show, such an expansion leads to more false detection of skin colors and reduces the effectiveness of skin segmentation as an attention-focus step in object detection tasks. Using lighting compensation techniques such as the one proposed by Hsu et al. [1] is probably a better approach to coping with extreme or biased lightings in skin detection.

The differences in the classification rates among the four chrominance feature vectors were more noticeable than the differences among the four 3-channel feature vectors. This is expected because the overlap between skin and nonskin alters when colors are projected from a 3D space to a 2D chrominance plane. The CbCr performed better compared to the other three chrominance feature vectors; this observation holds for histogram sizes of 256, 128, and 64 bins.

3.4 Analysis of Color Quantization

The Bayesian classifier with the histogram technique was again used to study the effects of color quantization on skin segmentation. In fact, for the Bayesian classifier, the level of color quantization is reflected in the histogram size used for pdf estimation. Clearly, a higher histogram size leads to finer pdf estimation but requires greater memory storage. Therefore, it is necessary to find a suitable level of color quantization in terms of segmentation accuracy and memory usage.

The ROC curves for the eight feature vectors across six histogram sizes are shown in Fig. 5. For all feature vectors, the

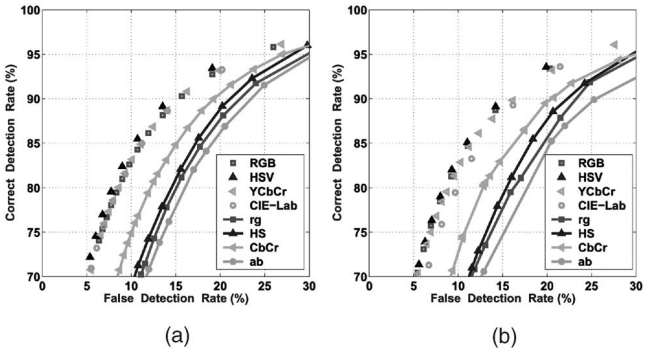


Fig. 4. ROC curves for different color representations. (a) 256 bins per channel. (b) 64 bins per channel.

best performance was found at the histogram size of 256 bins per channel. In general, larger histograms resulted in better ROC curves. However, the differences in classification rates for histogram sizes of 256, 128, and 64 bins per channel were quite small. This finding has a practical significance because for a 3D feature vector, reducing histogram size by half means reducing the memory storage by eight fold. The classification rates decreased sharply as the histogram size dropped below 32; chrominance feature vectors were the worst affected. Compared to other color spaces, the RGB and HSV color spaces were more robust to changes in the histogram size.

Our finding that larger histogram sizes tend to perform better is different from that of Jones and Rehg [4], who found that the histogram size of 32 bins per channel gave the best performance. We suspect that this could be caused by the difference in the training data used in our work and in Jones and Rehg's work. Intuitively, we know that if training samples are not sufficient, a larger histogram size will result in a noisier pdf estimate and a worse performance compared to a smaller histogram size. However, if training samples are sufficient, a larger histogram size will lead to a finer pdf estimate and hence better performance. To verify this hypothesis, we constructed 14 different Bayesian classifiers with histogram sizes of 256 and 32 bins per channel, using seven reduced training sets. The reduced training sets consisted of $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, $\frac{1}{32}$, $\frac{1}{64}$, and $\frac{1}{128}$ of the original training set, obtained through random subsampling. The classifiers were run on the test set in Table 3. We found that the 256-bin histogram size was more sensitive to the size of the training data compared to the 32-bin histogram size (see Fig. 6). Compared to the 32-bin histogram size, the 256-bin histogram size performed better for large training sets (containing more than $\frac{1}{8}$ of the original training set), and worse for small training sets (containing less than $\frac{1}{32}$ of the original training set).

TABLE 6
Classification Rates (CRs) of Eight Color Representations (Histogram Size = 256 Bins per Channel)

	All Channels				Only Chrominance Channels			
Color Representation	RGB	HSV	YCbCr	CIE-Lab	rg	HS	CbCr	ab
FDR = 10%	88.75	88.76	88.46	88.47	84.93	85.46	86.56	84.51
FDR = 15%	86.17	86.19	85.97	85.97	83.92	84.14	84.77	83.45
FDR = 20%	82.97	82.98	82.85	82.83	81.65	81.91	82.11	81.33
CR_{max}	89.79	89.80	89.41	89.43	84.95	85.58	86.75	84.52
99% confidence interval of CR_{max}	[89.74, 89.84]	[89.75 89.85]	[89.36 89.46]	[89.38 89.48]	[84.90 85.00]	[85.53 85.63]	[86.70 86.80]	[84.47 84.57]

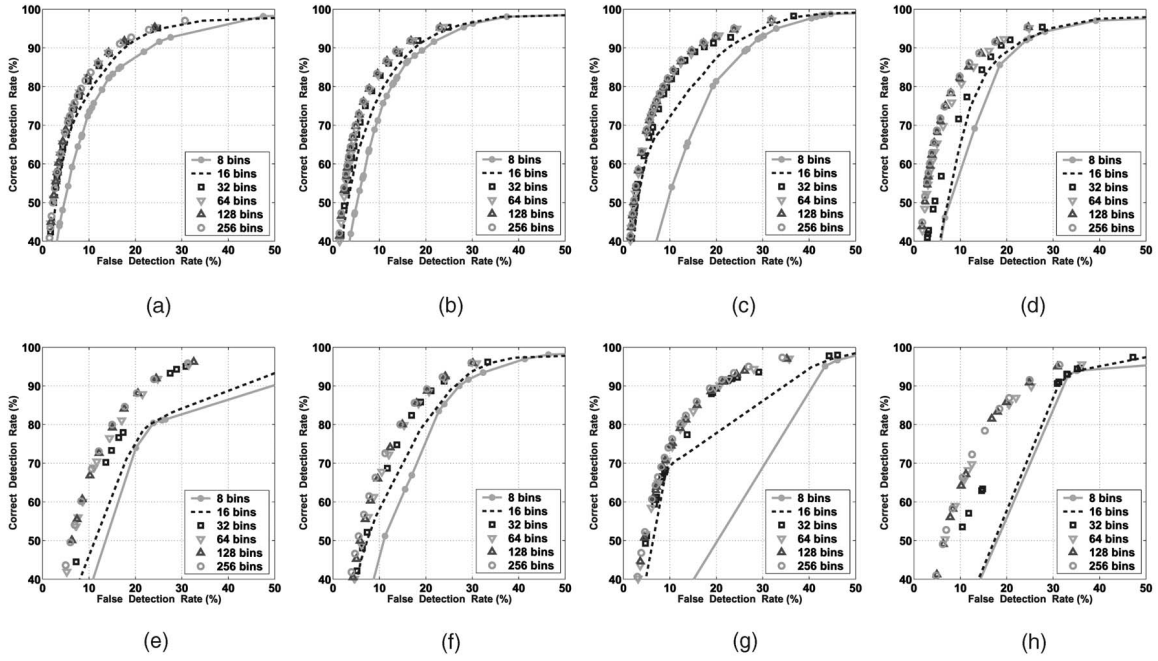


Fig. 5. ROC curves for different histogram sizes. (a) RGB. (b) HSV. (c) YCbCr. (d) CIE-Lab. (e) Normalized rg. (f) HS. (g) CbCr. (h) ab.

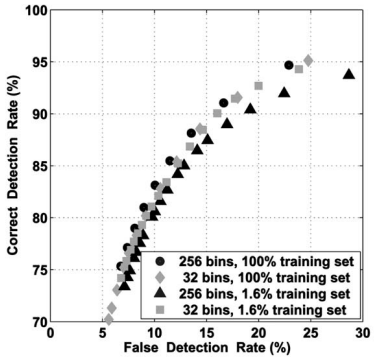


Fig. 6. ROC curves for different sizes of the training set.

4 Conclusions

An analysis of the pixelwise skin segmentation approach that uses color pixel classification is presented. The Bayesian classifier with the histogram technique and the multilayer perceptron classifier are found to have higher classification rates compared to other tested classifiers, including piecewise linear and Gaussian classifiers. The Bayesian classifier with the histogram technique is feasible for the skin color pixel classification problem because the feature vector has a low dimension and a large training set can be collected. However, the Bayesian classifier requires significantly more memory compared to the MLP and other classifiers. In terms of color representation, our study based on the Bayesian classifier shows that pixelwise skin segmentation is largely unaffected by the choice of color space. However, segmentation performance degrades if only chrominance channels are used and there are significant performance variations between different choices of chrominance. In terms of color quantization, we find that finer color quantization (a larger histogram size) gives better segmentation results. However, color pdf estimation can be done using histogram sizes as low as 64 bins per channel, provided that a large and representative training data set is used.

Acknowledgments

The authors would like to thank Tin Yuen Ke and Fok Hing Chi Tivive for taking part in developing the ECU database and the anonymous reviewers for their constructive comments. This research is supported in part by the Australian Research Council.

References

[1] R.-L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face Detection in Color Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696-707, May 2002.
[2] J. Yang and A. Waibel, "A Real-Time Face Tracker," *Proc. IEEE Workshop Applications of Computer Vision*, pp. 142-147, Dec. 1996.
[3] X. Zhu, J. Yang, and A. Waibel, "Segmenting Hands of Arbitrary Color," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 446-453, Mar. 2000.
[4] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," *Int'l J. Computer Vision*, vol. 46, no. 1, pp. 81-96, Jan. 2002.
[5] J. Brand and J. Mason, "A Comparative Assessment of Three Approaches to Pixel-Level Human Skin Detection," *Proc. IEEE Int'l Conf. Pattern Recognition*, vol. 1, pp. 1056-1059, Sept. 2000.
[6] J.-C. Terrillon, M.N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 54-61, Mar. 2000.
[7] M.C. Shin, K.I. Chang, and L.V. Tsap, "Does Colorspace Transformation Make Any Difference on Skin Detection?" *Proc. IEEE Workshop Applications of Computer Vision*, pp. 275-279, Dec. 2002.
[8] J.D. Foley, A.v. Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice.* New York: Addison Wesley, 1990.
[9] D. Chai and K N. Ngan, "Face Segmentation Using Skin Color Map in Videophone Applications," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551-564, 1999.
[10] C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 264-277, 1999.
[11] K. Sobottka and I. Pitas, "A Novel Method for Automatic Face Segmentation, Facial Feature Extraction and Tracking," *Signal Processing: Image Comm.*, vol. 12, no. 3, pp. 263-281, 1998.
[12] H. Wang and S.F. Chang, "A Highly Efficient System for Automatic Face Detection in Mpeg Video," *IEEE Trans. Circuits and Systems for Video Technology* vol. 7, no. 4, pp. 615-628 1997.
[13] H. Greenspan, J. Goldberger, and I. Eshet, "Mixture Model for Face Color Modeling and Segmentation," *Pattern Recognition Letters*, vol. 22, pp. 1525-1536, Sept. 2001.
[14] B. Menser and M. Wien, "Segmentation and Tracking of Facial Regions in Color Image Sequences," *SPIE Visual Comm. and Image Processing 2000*, vol. 4067, pp. 731-740, June 2000.
[15] M.-H. Yang and N. Ahuja, "Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Databases," *SPIE Storage and Retrieval for Image and Video Databases*, vol. 3656, pp. 45-466, Jan. 1999.
[16] S.L. Phung, D. Chai, and A. Bouzerdoum, "A Universal and Robust Human Skin Color Model Using Neural Networks," *Proc. INNS-IEEE Int'l Joint Conf. Neural Networks*, vol. 4, pp. 2844-2849, July 2001.
[17] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification.* John Wiley and Sons, 2001.
[18] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Upper-saddle, N.J.: Prentice-Hall, 1999.
[19] M. Fleck, D. Forsyth, and C. Bregler, "Finding Naked People," *Proc. European Conf. Computer Vision*, vol. 2, pp. 592-602 Apr. 1996.