Skin-Color Modeling and Adaptation

Jie Yang, Weier Lu, Alex Waibel

May 1997

CMU–CS–97–146

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

This research was sponsored by the Department of the Navy, Office of Naval Research under Grant No. N00014-93-1-0806.

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

Keywords: Skin color, color modeling, goodness-of-fit test, color adaptation, maximum likelihood

Contents

1 Introduction **1**

2 Skin-Color Distributions **3**
2.1 Color Space 4
2.2 Skin-Color Representation 5

3 Quantitative Analysis and Goodness-of-Fit Test **6**
3.1 Data Analysis 6
3.2 Goodness-of-fit Tests 9

4 Maximum Likelihood Adaptation **12**
4.1 Mean Vector Adaptation 15
4.2 Mean Vector and Covariance Matrix Adaptation 15
4.3 Applications 18

5 Conclusions **18**

List of Figures

1 An example of a face image and the skin-color occurrences in the RGB space 3
2 Skin-color distribution of the image in Figure 1 in the normalized color space 5
3 This figures shows how the face color data is collected from the original digital image. 7
4 This 3-D scatterplot displays all the colors that are present in 48 faces, which are taken as subsets from the original images. 7
5 These scatterplot displays all the colors that are present in one face. 8
6 The marginal q-q plot for the face colors of three different races. 11
7 The χ^2 test for the bivariate data of the face color for three races. 12
8 An example of face segmentation by skin-color model 13

Abstract

This report studies a statistical skin-color model and its adaptation. By quantitative analysis and goodness-of-fit test, we reveal that (1) skin-color differences among people can be reduced by intensity normalization, and (2) under a certain lighting condition, a skin-color distribution can be characterized by a multivariate normal distribution in the normalized color space. We then propose an adaptive model to characterize human skin-color distributions for locating human faces under different lighting conditions. The parameters of the model are adapted by a linear combination of the known parameters. The maximum likelihood criterion has been used to obtain the optimal estimation of the coefficients. The model has been successfully applied to a real-time face tracker and other applications.

1 Introduction

Human face perception is currently an active research area in the computer vision community. Locating and tracking human faces is a prerequisite for face recognition and/or facial expressions analysis, although it is often assumed that a normalized face image is available. In order to locate a human face, the system needs to capture an image using a camera and a framegrabber, to process the image, to search the image for important features, and then to use these features to determine the location of the face. In order to track a human face, the system not only needs to locate a face, but also needs to find the same face in a sequence of images.

Several systems of locating human face have been reported. Eigenfaces, obtained by performing a principal component analysis on a set of faces, have been used to identify faces [1]. By moving a window covering a subimage over the entire image, faces can be located within the entire image. [2] reports a face detection system based on clustering techniques. The system passes a small window over all portions of the image, and determines whether a face exists in each window. A similar system with better results has been claimed by [3]. A different approach for locating and tracking faces using skin-colors is described in [4, 5, 6].

Facial features, such as eyes, nose and mouth, are natural candidates for locating human faces. These features, however, may change from time to time. Occlusion and non-rigidity are basic problems with these features. For example, when a person rotates his/her head, depth changes can warp or occlude facial features. If we take a sequence of images of a person's rotating his/her head from left to right, the facial features will change as follows: in moving from a left sided face to a front face, the image of the left eye warps and the right ear appears (the inverse of occlusion); in moving from a front face to a right sided face, the left ear disappears (occlusion) and the image of the right eye warps. Four basic techniques are commonly used for dealing with feature variations: correlation templates [8, 9], deformable templates [10], spatial image invariants [11], and neural networks [2, 3]. These methods are, however, computational expensive and hardly achieve real-time performance. For example, the system described in [12] tracks object at about 5 frames/second speed with a 189 x 144 image by using a neural network to detect faces.

Color is another feature on human faces. Using skin-color as a feature for tracking a face has several advantages. Processing color is much faster than processing other facial features. Under certain lighting conditions, color is orientation invariant. This property makes motion estimation much easier because only a translation model is needed for motion estimation. However, color is not a physical phenomenon. It is a perceptual phenomenon that is related to the spectral characteristics of electro-magnetic radiation in the visible wavelengths striking the retina [13]. Tracking human faces using color as a feature has several problems. First, the color representation of a face obtained by a camera is influenced by many factors such as ambient light, object movement, etc. Second, different cameras produce significantly different color values even for the same person under the same lighting condition. Finally, human skin colors differ from person to person. In order to use color as a feature for face tracking, we have to solve these problems.

Much research has been directed to understanding and making use of color information. Color has been long used for recognition and segmentation [14, 15, 16, 17] and recently has been successfully used for road tracking [18] and face locating and tracking [4, 5]. Yang and Waibel proposed to use a statistical skin-color model for tracking human faces in real-time and developed a real-time face tracker achieved a rate of 30+ frames/second [6, 7]. While the system is successful, the skin-color model has yet to find a more rigorous theoretical foundation and quantitative justification. The general procedure for developing a distribution model includes finding cluster, extracting features (dimensionality reduction), and determining a distribution. In this report, we quantitatively investigate human skin color distributions. We demonstrate that:

- human skin-colors are clustered in the color space
- skin-color differences among people can be reduced by intensity normalization
- under a certain lighting condition, a skin-color distribution can be characterized by a multivariate normal distribution in the normalized color space.

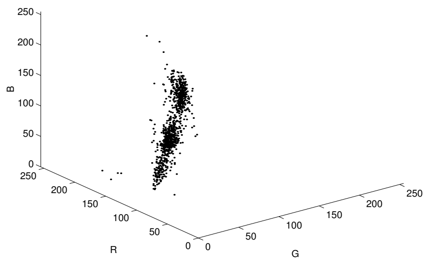
A common believe is that different people have different color appearances. This study shows that such a difference lies largely in intensity than color itself. By color normalization, the skin-color difference among different people can be greatly reduced. Furthermore, using goodness-of-fit techniques, we verify that under a certain lighting condition, a human skin-color distribution is a normal distribution. Based on these results, we present an adaptive parametric model to characterize human skin-color distributions for different people under different lighting conditions. Since a linear transformation of a normal distribution is still a normal distribution, the different skin-color distributions can be considered as transformed distributions from other distributions. We propose to use a linear combination of the known parameters to predict or approximate new parameters. The maximum likelihood method has been used to estimate the coefficients of the linear transformation. We investigate two cases: estimating mean vector only and estimating both mean vector and covariance matrix. We derive the maximum likelihood estimates for both cases.

The remainder of the report is structured as follows. Section 2 discusses the general problem of skin-color distributions. Section 3 performs quantitative analysis and goodness-of-fit test on skin-color distributions. Section 4 addresses the maximum likelihood adaptation of skin-color model to different lighting conditions and different people. We close with a discussion of future work.

2 Skin-Color Distributions

Color is the perceptual result of light in the visible region of the spectrum, having wavelengths in the region of 400 nm to 700 nm, incident upon the retina. Physical power (or radiance) is expressed in a spectral power distribution. A color histogram is a distribution of colors in the color space and has long been used by the computer vision community in image understanding. For example, analysis of color histograms has been a key tool in applying physics-based models to computer vision. It has been shown that color histograms are stable object representations unaffected by occlusion and changes in view, and that they can be used to differentiate among a large number of objects [16]. In the mid-1980s, it was recognized that the color histogram for a single inhomogeneous surface with highlights will have a planar distribution in color space [19]. It has since been shown that the colors do not fall randomly in a plane, but form clusters at specific points [20, 21]. The color histograms of human skin coincide with these observations. The Figure 1 shows a face image and the skin-color occurrences in the RGB color space (256x256x256). The skin-colors are clustered in a small area in the RGB color space, i.e., only a few of all possible colors actually occur in a human face.

(a) Face image (color!)



(b) Skin-color occurrences

Figure 1: An example of a face image and the skin-color occurrences in the RGB space

2.1 Color Space

A variety of spectral distributions of light can produce perceptions of color which are indistinguishable from one another. The human retina has three different types of color photoreceptor cone cells, which respond to incident radiation with somewhat different spectral response curves. Based on the human color perceptual system, three numerical components are necessary and sufficient to describe a color, provided that appropriate spectral weighting functions are used. Theoretically, color coordinates can be defined as product integrals of the stimulus spectrum $U(n)$ with three linearly independent color matching functions $\bar{r}(\nu)$, $\bar{g}(\nu)$, $\bar{b}(\nu)$,

$$R = \int_{\nu_1}^{\nu_2} \bar{r}(\nu)U(\nu)d\nu, \tag{1}$$

$$G = \int_{\nu_1}^{\nu_2} \bar{g}(\nu)U(\nu)d\nu, \tag{2}$$

$$B = \int_{\nu_1}^{\nu_2} \bar{b}(\nu)U(\nu)d\nu. \tag{3}$$

where ν is the frequency of the light stimulus.

It is well known that different people have different skin-color appearances. Even for the same person, his/her skin-color appearance will be different if he/she wears different clothes or under different lighting conditions. In other words, many factors contribute to human skin-color appearance. In order to characterize skin-color, we hope to find a color space in which skin-colors are less variants. For human color perception, a 3D color space such as an RGB space, is essential. Most video cameras use an RGB model; other color models can be easily converted into an RGB model. However, an RGB space is not necessarily essential for all other problems. In the problem of locating human faces, intensity is not important. Therefor we can remove it from the original information by normalization. Our experiments reveal that human color appearances differ more in brightness than in color itself. If we can remove the brightness from the color representation, the difference among human skin-colors can be greatly reduced.

The human visual system adapts to different brightness and various illumination sources such that a perception of color constancy is maintained within a wide range of environmental lighting conditions [22]. Therefore it is possible for us to remove brightness from the skin-color representation while preserving an accurate but low dimensional color information. In fact, a triple $[r, g, b]$ in the RGB space represents not only color but also brightness. If the corresponding elements in two points, $[r_1, g_1, b_1]$ and $[r_2, g_2, b_2]$, are proportional, i.e.,

$$\frac{r_1}{r_2} = \frac{g_1}{g_2} = \frac{b_1}{b_2}, \tag{4}$$

they have the same color but different brightness. The brightness can be removed from color space by normalization. Chromatic colors (r, g) [13], known as "pure" colors in the absence of brightness, are defined by a normalization process:

$$r = \frac{R}{(R+G+B)}, \tag{5}$$

$$g = \frac{G}{(R+G+B)}. \tag{6}$$

In fact, (5) and (6) define an $R^3 \rightarrow R^2$ mapping. Color blue is redundant after the normalization because $r + g + b = 1$. It has been showed that the differences of the color distributions have been reduced after the normalization. In other words, skin-colors of different people are less variant in the normalized color space. This result is significant because it provides evidence of the possibility of modeling human faces with different color appearances in the chromatic color space.

2.2 Skin-Color Representation

We have so far revealed that human skin-colors cluster in the color space and are less variant in the chromatic color space. We are further interested in the representation of the skin-color distributions. However, the skin-color distribution is related not only to the skin-color, but also to the illumination color because only those colors can be reflected. For example, sunlight will shift color histograms towards blue because it contains more blue than fluorescent lighting. Therefore, it is impossible to characterize all the skin-color distributions using a fixed model. On the other hand, although skin colors of different people appear to vary over a wide range, it is possible to model the skin-color distribution of each individual under a certain lighting condition. Since the skin-color distribution has only two variables in the normalized color space, it is convenient to investigate it graphically. Figure 2 shows a skin color distribution of the image in Figure 1.

(a) Global view (b) Local view

Figure 2: Skin-color distribution of the image in Figure 1 in the normalized color space

Moreover, we have found that the shape of the skin-color distribution of a person remains similar although there is a shift in the distribution under changing lighting conditions. By closely investigating the face color cluster, we have discovered that the distribution has a regular shape. By comparing the shape of skin-color distributions with a bivariate normal distribution, it concludes that it is possible to use a bivariate normal distribution to characterize the skin-color distributions

3 Quantitative Analysis and Goodness-of-Fit Test

We present in this section the human face color data along with the quantitative analysis to determine its statistical distribution using goodness-of-fit techniques. We demonstrate that the composition of human skin-color distributions can be approximated by bivariate normal distributions as we have asserted earlier in the previous section. The data we used in this study are from a large pool (about one thousand) of color digital images collected from the public domain on the internet as well as those recorded in our multimedia lab. A large portion of our face database were down-loaded from *http://pics.psych.stir.ac.uk/*, which contains a collection of images for use in psychology and visual science research. We choose approximately seven hundred color images from this database that cover both both genders, a wide age range, and various lighting conditions. To compliment the shortcomings of these data, we also built a database in our own lab which contains facial images of people of the Caucasian, African American, and Asian races and of both genders, and the lighting intensity were varied to cover the most normal application conditions.

3.1 Data Analysis

We first investigate the problem of color space for representing human skin-colors. A digital color image is actually a two-dimensional array of pixels with a finite size, each of which is specified by a set of intensities for three independent colors, usually red, green, and blue. Although the three numerical values for the image coding could, in theory, be provided by a color specification system, a practical image coding system needs to be computationally efficient and cannot afford unlimited precision. In this work, we represent color in the RGB color space with 8 bits for each color band, i.e., there are 256^3 bins into which a pixel may fall. Thus, a particular color is conveniently represented by a point in the color space whose axes corresponds to the intensity levels of each color: Red, Green, and Blue (see Figure 4). For each image, the sample data is collected from a region occupied mainly by the human skin, such as a subset frame shown in Figure 3.

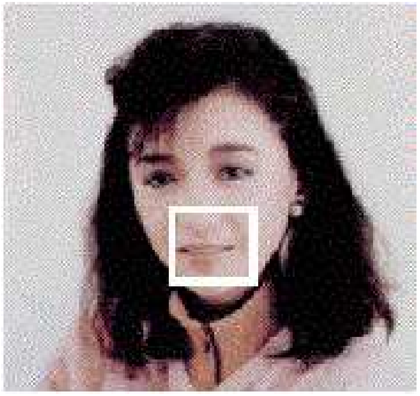
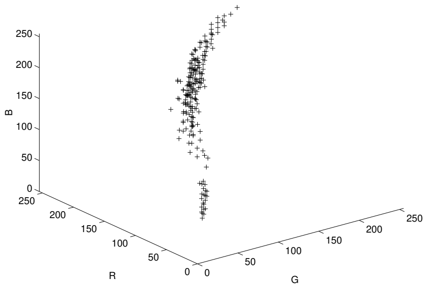


Figure 3: This figures shows how the face color data is collected from the original digital image.



3-D scatterplot

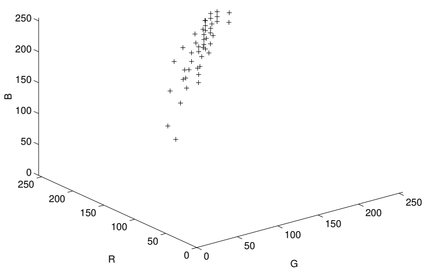
Figure 4: This 3-D scatterplot displays all the colors that are present in 48 faces, which are taken as subsets from the original images.

Figure 4 shows in the RGB space a typical aggregated color occurrence distribution from a random set of 48 human faces of various race, age, gender, and lighting conditions. Each point in the figure designates the presence of a particular color as specified by the corresponding coordinate values. The total number of images included in such a set is limited by the memory resource of the system associated with the statistical analysis software we used, and we did not attempt to migrate the computation to a more powerful machine because through experimentation, we found images beyond 20 adds little to the aggregated color pool. This attribute is further affirmed when we also analyzed several similar random sets of images and found no qualitative differences are found among them. As an example, the means and variances of each color for the set shown in Figure 4 are

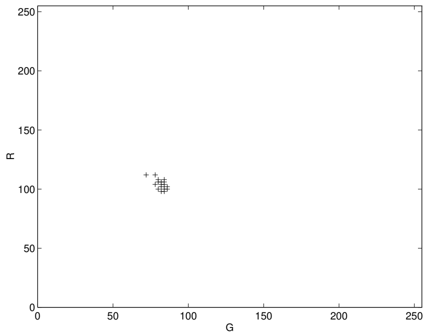
$$
\begin{aligned}
m_{green} &= 142.9157, \\
m_{red} &= 188.9069, \\
m_{blue} &= 115.1863; \\
s_{green} &= 45.3306, \\
s_{red} &= 58.3542, \\
s_{blue} &= 43.397.
\end{aligned}
$$

It is evident from these results that the color occurrences of all human faces under various lighting conditions (at least for all what we have collected, in which we carefully tried to cover most conceivable application scenarios) are well confined in a specific region in the color space. This property, as we have confirmed, is the foundation of skin-color modeling.

Compared to the aggregated color distribution, colors from individual faces are expectedly more narrowly distributed, as shown in Figure 5. The corresponding means and variances



(a) Single face 3-D scatterplot



(a) Single face 2-D scatterplot

Figure 5: These scatterplot displays all the colors that are present in one face.

of three examples are

$$
\begin{aligned}
m_{green} &= 185.7177, \\
m_{red} &= 234.2947, \\
m_{blue} &= 151.1090; \\
s_{green} &= 30.4088, \\
s_{red} &= 26.7735, \\
s_{blue} &= 25.6779.
\end{aligned}
$$

Different people have different color appearances. This raises a question: can we reduce such differences in some way? While the RGB space represent the true color of the image, it is not necessarily the best space for characterizing skin-colors. We hope to find a color space where the skin-colors are less variant. The skin-color appearance is related not only to the skin-color, but also to the illumination color. We want to minimize the effects from the illumination. An efficient way is normalization by Equation(5) and Equation(6). There is a two-fold benefit from this R^3 to R^2 mapping. First, it reduces the number of parameters needed for modeling skin-colors, resulting in a much less complex system. Second, the mapping also reduces the variances of skin-color distributions, as is obvious from the comparison of the means and variance data for the 3-D case to the 2-D reduction:

$$\begin{aligned} m_{green} &= 81.5879 \\ m_{red} &= 104.2225 \\ s_{green} &= 3.8858 \\ s_{red} &= 4.9317. \end{aligned}$$

(computed from the same data as shown in Figures 5). These attributes are essential to system performance and robustness.

3.2 Goodness-of-fit Tests

The remaining challenge is to determine what statistical distribution function best describes the data. We have observed that the skin-color distributions are Gaussian-like distributions. Unlike most of the methods used in engineering statistics assume a normal distribution of the measured data, we will examine whether the measured data of a sample do indeed have a normal distribution by goodness-of-fit techniques. Goodness-of-fit tests test the conformity of the observed data's empirical distribution function with a posited theoretical distribution function.

Thus, we have a *NULL hypothesis:*
human skin-color is normally distributed in a normalized bivariate space.

Our task is to determine whether or not we can build up enough evidence to reject the hypothesis. We tested the hypothesis with more than a thousand images using the goodness-of-fit method, which is a widely used tool in the confirmatory statistical analysis that we need to accomplish.

An immediate difficulty of the task is that there is no commonly agreed analytical tool available to test the normality of a bivariate distribution [25]. Since the marginal distribution of a bivariate normal must also be normal, it would be efficient that we test for marginal distributions first. If the test fails at this level, we would know that the data cannot be bivariately normal and the NULL hypothesis should be rejected. It would thus save us the trouble of further bivariate level tests.

To perform marginal normality test, we deal with each variable separately as if there is no other variables. For such one dimensional normality test, a few goodness-of-fit techniques exist, and we employ here the most popularly used *Quantile-Quantile* plot (or Q-Q plot) graphical technique due to its straightforwardness and simplicity. Given a set of n sample data, the *quantiles* are the same data ordered from the smallest to largest. Corresponding to each of the data point, its order position (e.g. ith) in that data set is associated with a cumulative percentage (called p value) in the occurrence distribution of that data $(i - 0.5)/n$ (the interested readers are referred to textbooks on the subject, e.g. [25].)

Once the data are ordered in place and with each p value calculated, the corresponding variable value of an ideal normal distribution can be computed by numerically solving the inverse function of the cumulative normal distribution, whose mean and variance are estimated from the data. Hence a one-to-one match between the test data and theoretical data can be constructed. These matches can be plotted together on a standard normal distribution scale. The deducted normal distribution would show up as a straight line with its intercept and slope valued according to the group's mean and variance, and if the tested data is indeed normally distributed, the data points should basically match the line.

Plots in Figures 6 are the marginal Q-Q plots of Asian, African American, and Caucasian races, respectively. Because of the limitation of the space, we show here only an extremely small selection of our results that capture the major features of the several hundred plots we have produced during this analysis. While the intercepts and slopes of the lines are different (expectedly, for different people and lighting condition) from plot to plot, all the plots have most of the data points fall on or scattered closely nearby these lines.

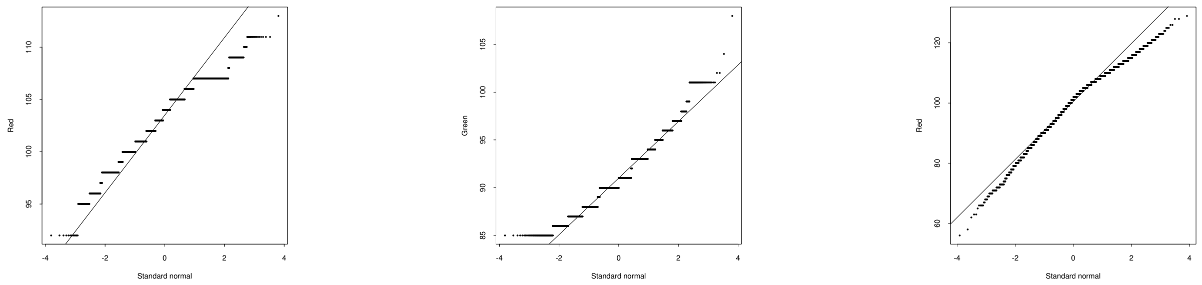
These figures demonstrate that our data can be said to be at least marginally normal, and thus we have failed to reject the NULL hypothesis that states that the face color data satisfies the normal distribution by marginal normality tests.

Having passed the marginal tests, we need to further verify the bivariate normality. A relatively simple and effective way to determine deviations from the normal distribution is the 2-dimensional Quantile-Quantile plot (Q-Q plot) method. It is based on the fact that, for normally distributed multivariate vector data $\mathbf{x}$ of dimension n, the transformation [24]

$$(\mathbf{x} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu) = \mathbf{u}'\mathbf{u} \tag{7}$$

where μ is the mean and Σ the dispersion matrix:

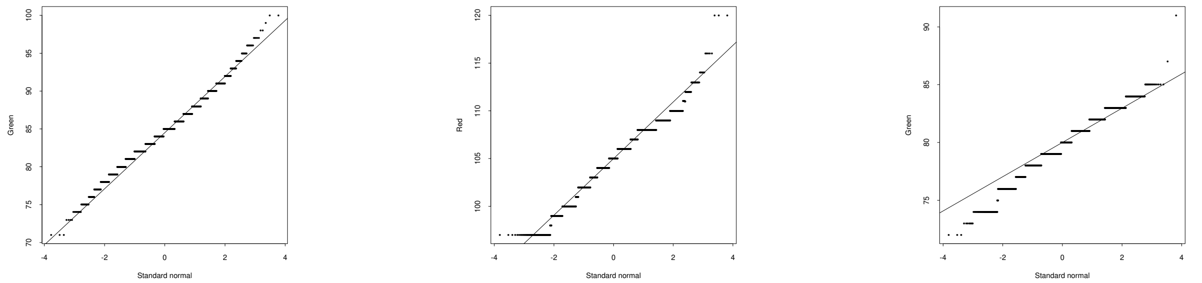
$$\mathbf{\Sigma} = \mathbf{E}\left[\{\mathbf{x} - \mathbf{E}(\mathbf{x})\}\{\mathbf{x} - \mathbf{E}(\mathbf{x})\}'\right].$$



(a) Red data of an Asian

(b) Green data of an Asian

(c) Red of an African American



(d) Green of an African American

(e) Red of a Caucasian

(f) Green of a Caucasian

Figure 6: The marginal q-q plot for the face colors of three different races.

results in the square of a standard normal $\mathbf{u}$, i.e. $u \sim \mathbf{N}(0, I_n)$. Since $u'u = \sum_{i=1}^{n} u_i^2$ is the sum of n independent N(0,1) variates, therefore

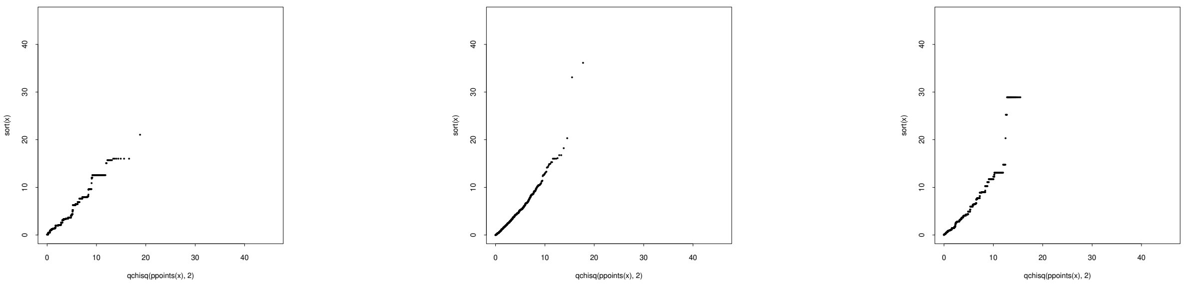
$$\mathbf{z} = (\mathbf{x} - \mu)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu) = \mu'\mu \tag{8}$$

has a χ^2 distribution with n degrees of freedom. We can thus test the normality of $\mathbf{x}$ by testing z_i's to the χ_n^2 distribution.

The graphic testing procedure is as the following: similar to the marginal normality testing procedure, we first calculate z_i's and sort them in the increasing order. Corresponding to each of the data point, its order position (e.g. ith) in that data set is associated with a cumulative probability (called p value) of $(i - 0.5)/n$ (e.g. [25]).

The quantile of the distribution is computed by numerically solving the inverse function of the cumulative distribution function at each specified probability point. Thus if z_i's are truly independent observations from a χ_n^2 distribution (in our case $n = 2$), then the plot of z_i against the χ_n^2 quantiles should yield a straight line.

Plots in Figure 7 shows the Q-Q plot against χ^2 distributions of 2 degree of freedom for face colors of Asian, African, and Caucasian races respectively. The result does not display any evidence of a significant deviation of our data from the line, and we can safely assert



(a) Asian (b) African American (c) Caucasian

Figure 7: The χ^2 test for the bivariate data of the face color for three races.

that our tests failed to reject the NULL hypothesis; thus normality can be assumed for the any formal applications and analysis of the data.

Having confirmed that our color data basically follows a normal distribution, however, we also can notice that the deviation from the straight line is sometimes significant, especially at the lower end of the line. These phenomena indicates there are some outliers in the data. The reason, in our belief, is partly contributed to by the rounding of the color values to an 8-bit integer (between 0 and 255), and it may be verified only by increase the the number of bits for each color. Another major cause of deviation from normality is the cosmetic makeups on the faces, which makes a particular color predominantly strong and the thus the distribution would have a diminishing "shoulder."

4 Maximum Likelihood Adaptation

We have verified that under a certain lighting condition human skin-colors can be characterized by a multivariate normal distribution, i.e., $N(\mu, \Sigma)$, where $\mu = (r, g)$ with

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i, \quad \bar{g} = \frac{1}{N}\sum_{i=1}^{N} g_i, \tag{9}$$

and

$$\Sigma = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix}. \tag{10}$$

A direct application of the skin-color model is to locate a face in an image. A straightforward way to locate a face is to match the model with the input image to find the face color clusters. Each pixel of the original image is converted into the chromatic color space and then compared with the distribution of the skin-color model. Since the skin colors occur in a small area of the chromatic color space, the matching process is very fast. This is useful for real-time face tracking. Figure 8 b is an example of color segmentation for Figure 8 a image using the skin-color model. Pixels with a high gray-scale value in Figure 8 correspond to frequently occurring skin colors. Although the skin-color region contains the eyes and the lips as well as background, there is little difficulty to locate a face based on the result of Figure 8.

(a) Original image

(b) Segmented face

Figure 8: An example of face segmentation by skin-color model

Although under a certain environment the skin-color distribution of each individual is a multivariate normal distribution, the parameters of the distribution for different people and different lighting conditions are different. A number of viewing factors, such as light sources, background colors, luminance levels, and media, impact greatly on the change in color appearance of an image. Most color-based systems are sensitive to changes in viewing environment. Even under the same lighting conditions, background colors such as colored clothes may influence skin-color appearance. Furthermore, if a person is moving, the apparent skin-colors change as the person's position relative to camera or light changes. Therefore, the ability of handling lighting changes is the key to success for a skin-color model.

There are two schools of philosophy to handle environment changes: tolerating and adapting. Color constancy refers to the ability to identify a surface as having the same color under considerably different viewing conditions. Although human beings have such ability, the underlying mechanism is still unclear. A few color constancy theories have demonstrated success on real images [23]. On the other hand, the adaptive approach provides an alternative to make a color model useful in a large range. Instead of emphasizing the recovery of the spectral properties of light sources and surfaces that combine to produce the reflected lights, the goal of adaptation is to transform the previously developed color model to the new environment.

In this report we present a method to adapt the skin-color model. The basic idea is to use a linear combination of the known parameters to predict the new parameters. Suppose that $\mathbf{X}$ has a multivariate normal distribution. If $\mathbf{Y} = \mathbf{BX}$ is any linear transformation of

$\mathbf{X}$, where $\mathbf{B}$ is an $(m \times p)$ matrix of real numbers with $m \leq p$ and rank m, then $\mathbf{Y}$ also has a multivariate normal distribution. Based on the identification of the skin-color distribution at each sampling point, we can obtain its mean vector and covariance matrix. We can consider this set of parameters is a linear combination of the past r sets of parameters, such that

$$\hat{\mu} = \sum_{k=1}^{r} \alpha_k \mathbf{m_k}, \tag{11}$$

where $\hat{\mu}$ is the estimated mean vector; $\alpha_i \leq 1$, $k = 1, \ldots, r$, are weighting factors; $\mathbf{m_k}$, $k = 1, \ldots, r$, are the previous mean vectors.

$$\hat{\Sigma} = \sum_{k=1}^{r} \beta_k S_k, \tag{12}$$

where $\hat{\Sigma}$ is the estimated covariance matrix; $\beta_k \leq 1$, $k = 1, \ldots, r$, are weighting factors; S_k, $k = 1, \ldots, r$ are the previous covariance matrices.

The weighting factors in (11) and (12) determine how much the past parameters will influence current parameters. We then use this set of coefficients to predict the new parameters. We will use the maximum likelihood criterion to find the best set of coefficients for the prediction. Since we have verified that the skin-color distribution is a normal distribution, The likelihood function of N observations on $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2})$ in the normalized bivariate color space is

$$L = \frac{1}{(2\pi)^N \left|\Sigma\right|^{\frac{1}{2}N}} exp\left[-\frac{1}{2}\sum_{k=1}^{N}(\mathbf{x_k} - \mu)'\mathbf{\Sigma^{-1}}(\mathbf{x_k} - \mu)\right]. \tag{13}$$

The logarithm of the likelihood function is

$$\log L = -N\log(2\pi) - \frac{1}{2}N\log\left|\Sigma^{-1}\right| - \frac{1}{2}\sum_{k=1}^{N}(\mathbf{x_k} - \mu)'\mathbf{\Sigma^{-1}}(\mathbf{x_k} - \mu). \tag{14}$$

Since $\log L$ is an increasing function of L, its maximum is at the same point in the space of μ as the maximum of L.

Let the sample mean and variance be

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x_k} = \begin{pmatrix} \frac{1}{N}\sum_{k=1}^{N} x_{1k} \\ \frac{1}{N}\sum_{k=1}^{N} x_{2k} \end{pmatrix} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}. \tag{15}$$

$$C = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x_k} - \bar{\mathbf{x}})(\mathbf{x_k} - \bar{\mathbf{x}})' \tag{16}$$

We have

$$\sum_{k=1}^{N}(\mathbf{x_k}-\mu)(\mathbf{x_k}-\mu)' = NC + N(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'. \tag{17}$$

Using this result and the properties of the trace of a matrix we can rewrite (14) as

$$\log L = -N\log(2\pi) - \frac{1}{2}N\log\left|\Sigma^{-1}\right| - \frac{1}{2}Ntr\ \Sigma^{-1}C - \frac{1}{2}N(\bar{\mathbf{x}}-\mu)'\mathbf{\Sigma^{-1}}(\bar{\mathbf{x}}-\mu). \tag{18}$$

We will use (18) to derive the maximum likelihood equations. We will discuss two cases: (1) adapting mean vector only; and (2) adapting both mean vector and covariance matrix.

4.1 Mean Vector Adaptation

In this case, the covariance matrix is assumed to be a constant and the mean vector μ is assumed to be a linear combination of the previous mean vectors

$$\hat{\mu} = \sum_{k=1}^{r}\alpha_k\mathbf{m_k},$$

$$\hat{\Sigma} = \Sigma,$$

where $\hat{\mu}$ is the estimated mean vector; α_k, $k = 1,\ldots,r$, are unknown coefficients; $\mathbf{m_k}$, $k = 1,\ldots,r$, are the previous mean vectors; Σ is the covariance matrix.

By setting the derivatives of the likelihood function (18) with respect to $\alpha_k, k = 1,\ldots,r$, to 0, the equations for the maximum likelihood estimates are

$$\sum_{k=1}^{r}\mathbf{m_j}'\Sigma^{-1}\mathbf{m_k}\hat{\alpha}_j = \mathbf{m_j}'\Sigma^{-1}\bar{\mathbf{x}}, \quad j = 1,\ldots,r \tag{19}$$

If $\sum_{k=1}^{r}\mathbf{m_j}'\Sigma^{-1}\mathbf{m_k} \neq 0$, j= 1, ..., r, we have

$$\hat{\alpha}_j = (\sum_{k=1}^{N}\mathbf{m_j}'\Sigma^{-1}\mathbf{m_k})^{-1}\mathbf{m_j}'\Sigma^{-1}\bar{\mathbf{x}}, \quad j = 1,\ldots,r \tag{20}$$

4.2 Mean Vector and Covariance Matrix Adaptation

In this case, the mean vector is assumed to be a linear combination

$$\hat{\mu} = \sum_{k=1}^{r}\alpha_k\mathbf{m_k},$$

and the covariance matrix is assumed to be a linear combination

$$\hat{\Sigma} = \sum_{k=1}^{r}\beta_k S_k,$$

where $\hat{\mu}$ is the estimated mean vector; α_k and β_k, $k = 1, \ldots, r$, are unknown coefficients; $\mathbf{m_k}$, $k = 1, \ldots, r$, are the previous mean vectors; $\hat{\Sigma}$ is the estimated covariance matrix; and S_k, $k = 1, \ldots, r$ are the previous covariance matrices.

Since the two sets of estimates are asymptotically independent, each set of parameters can be estimated as when the other set of parameters is known. $\alpha_k, k = 1, \ldots, r$, can be estimated by (19). We will derive the maximum likelihood estimates for $\beta_k, k = 1, \ldots, r$, by the same likelihood function (18). Because Σ^{-1} is positive definite, (18) is maximized with respect to μ at $\hat{\mu} = \bar{\mathbf{x}}$. The logarithm of the reduced likelihood function is then proportional to

$$-\log(2\pi) - \log\left|\Sigma^{-1}\right| - tr\ \Sigma^{-1}C. \tag{21}$$

By differentiating (21) with respect to $\beta_k, k = 1, \ldots, r$, we have

$$\frac{\partial}{\partial \beta_j}\Sigma^{-1} = -\Sigma^{-1}S_j\Sigma^{-1}, \quad j = 1, \ldots, r, \tag{22}$$

and

$$\frac{\partial}{\partial \beta_j}\log|\Sigma| = tr\ \Sigma^{-1}S_j \quad j = 1, \ldots, r. \tag{23}$$

The derivatives of (21) are $-tr\ \Sigma^{-1}S_j + tr\ \Sigma^{-1}S_j\Sigma^{-1}C$, and the maximum likelihood estimate equations are

$$tr\ (\sum_{k=1}^{r}\hat{\beta}_k S_k)^{-1}S_j = tr\ (\sum_{k=1}^{r}\hat{\beta}_k S_k)^{-1}S_j(\sum_{k=1}^{r}\hat{\beta}_k S_k)^{-1}C, \quad j = 1, \ldots, r. \tag{24}$$

The maximum likelihood estimation problem for the multivariate normal distribution with linear structure of mean vector and covariance matrix has been studied by many researchers [26, 27, 28, 29, 30]. In general explicit solutions for the equation (24) do not exist and estimates must be performed by iterative numerical techniques. In the following we present an algorithm based on the estimate procedure proposed by Anderson [28].

The basic idea of the algorithm is to iteratively estimate two sets of parameters independently. In order to iteratively estimate $\hat{\alpha}_k^{(i)}$ and $\hat{\beta}_k^{(i)}$, where the superscript (i) denotes the ith iteration, we can rewrite (24) as

$$\sum_{k=1}^{r} tr\ (\hat{\Sigma}^{(i-1)})^{-1}S_j(\hat{\Sigma}^{(i-1)})^{-1}S_k\hat{\beta}_k^{(i)} = tr\ (\hat{\Sigma}^{(i-1)})^{-1}S_j(\hat{\Sigma}^{(i-1)})^{-1}C^{(i)}, \quad j = 1, \ldots, r, \tag{25}$$

The initial values of $\hat{\alpha}_k^{(0)}$ and $\hat{\beta}_k^{(0)}, k = 1, \ldots, r$ are obtained by setting $\hat{\Sigma}^{(0)}$ to identity matrix I. The iteration proceeds by estimating parameters in the order of $\hat{\alpha}_k, k = 1, \ldots, r$, $\hat{\mu}$, C, $\hat{\beta}_k, k = 1, \ldots, r$, and Σ. The iteration is terminated if $\hat{\alpha}_k^{(i)}, k = 1, \ldots, r$ do not differ significantly from $\hat{\alpha}_k^{(i-1)}, k = 1, \ldots, r$. The algorithm is as follows.

Algorithm

1. Initialization

$$\hat{\alpha}_j^{(0)} = (\sum_{k=1}^{r} \mathbf{m_j}'\mathbf{m_k})^{-1}\mathbf{m_j}'\bar{\mathbf{x}}, \quad j = 1, \ldots, r,$$

$$\hat{\mu_j}^{(0)} = \sum_{k=1}^{r} \hat{\alpha}_k^{(0)}\mathbf{m_k}, \quad j = 1, \ldots, r,$$

$$C^{(0)} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x_k} - \bar{\mathbf{x}})(\mathbf{x_k} - \bar{\mathbf{x}})' + (\mathbf{x_k} - \hat{\mu}^{(\mathbf{0})})(\mathbf{x_k} - \hat{\mu}^{(\mathbf{0})})'$$

$$\sum_{k=1}^{r} tr\ S_j S_k \hat{\beta_k}^{(0)} = tr\ S_j C^{(0)}, \quad j = 1, \ldots, r,$$

$$\hat{\Sigma}^{(0)} = \sum_{k=1}^{r} \hat{\beta_k}^{(0)} S_k,$$

2. Iteration

$$\hat{\alpha}_j^{(i)} = (\sum_{k=1}^{r} \mathbf{m_j}'(\hat{\Sigma}^{(i-1)})^{-1}\mathbf{m_k})^{-1}\mathbf{m_j}'(\hat{\Sigma}^{(i-1)})^{-1}\bar{\mathbf{x}}, \quad j = 1, \ldots, r,$$

$$\hat{\mu_j}^{(i)} = \sum_{k=1}^{r} \hat{\alpha}_k^{(i)}\mathbf{m_k}, \quad j = 1, \ldots, r,$$

$$C^{(i)} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x_k} - \bar{\mathbf{x}})(\mathbf{x_k} - \bar{\mathbf{x}})' + (\mathbf{x_k} - \hat{\mu}^{(\mathbf{i})})(\mathbf{x_k} - \hat{\mu}^{(\mathbf{i})})'$$

$$\sum_{k=1}^{r} tr\ (\hat{\Sigma}^{(i-1)})^{-1} S_j (\hat{\Sigma}^{(i-1)})^{-1} S_k \hat{\beta_k}^{(i)} = tr\ (\hat{\Sigma}^{(i-1)})^{-1} S_j (\hat{\Sigma}^{(i-1)})^{-1} C^{(i)}, \quad j = 1, \ldots, r,$$

$$\hat{\Sigma}^{(i)} = \sum_{k=1}^{r} \hat{\beta_k}^{(i)} S_k,$$

3. If $\max(|\beta_j^{(i)} - \beta_j^{(i-1)}|,\ j = 1, \ldots, r) \leq \epsilon$ for a small number $\epsilon > 0$, stop; otherwise goto step 2.

It has been shown that the solution of these estimation equations is asymptotically efficient provided that estimate of Σ is consistent [28].

4.3 Applications

The adaptive skin-color model has been applied to many applications. The model plays a key role in the real-time face tracker [6, 7]. The system has achieved a rate of 30+ frames/second with 305 x 229 input sequences of images on both HP and Alpha workstations. The system can track a person's face while the person walks, jumps, sits and rises. The QuickTime movies of demo sequences in different situations and on different subjects can be found in the web site http://www.is.cs.cmu.edu/. The skin-color model has also been applied to other applications such as tele-conferencing [31], gaze tracking [32], and lip-reading [33].

5 Conclusions

We have proposed a statistical skin-color model for tracking human faces in real-time. We have shown that differences of skin-color appearances of different people can be reduced by normalization by data analysis. Using goodness-of-fit techniques, we have further verified that the skin-color distribution of each individual under a certain lighting condition can be characterized by a multivariate normal distribution. Based on these results, we have proposed an adaptive skin-color model to characterize human faces different views under different lighting conditions. We have used a linear combination of the known parameters to predict or approximate new parameters. The maximum likelihood method has been used to estimate the coefficients of the linear transformation. We have investigated two cases: estimating mean vector only and estimating both mean vector and covariance matrix. In the later case, an iterative algorithm has been employed to obtain the optimal coefficients. The feasibility of the model has been demonstrated by a real-time face tracker and other applications in human computer interaction.

References

[1] M.A. Turk and A. Pentland, "Face recognition using eigenfaces," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591, Maui, HI, USA, 1991.

[2] K. Sung and T. Poggio, "Example-based learning for view-based human face detection," Technical Report 1521, MIT AI Lab, 1994.

[3] H.A. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes," Technical Report CMU-CS-95-158, CS department, CMU, 1995.

[4] M. Hunke and A. Waibel, "Face locating and tracking for human-computer interaction," Proc. Twenty-Eight Asilomar Conference on Signals, Systems & Computers, Monterey, CA, USA, 1994.

[5] T.C. Chang, T.S. Huang, and C. Novak, "Facial feature extraction from color images," Proc. the 12th IAPR International Conference on Pattern Recognition, Vol. 2, pp. 39-43, 1994.

[6] J. Yang and A. Waibel, "Tracking human faces in real-time," Technical Report CMU-CS-95-210, CS department, CMU, 1995.

[7] J. Yang and A. Waibel, "A real-time face tracker," Proceedings of the Third IEEE Workshop on Applications of Computer Vision (Sarasota, Florida, 1996), pp. 142-147.

[8] R. Brunelli and T. Poggio, "Face recognition: features versus templates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, pp. 1042-1052, Oct. 1993.

[9] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspace for face recognition," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 84-91, Seattle, WA, USA, 1994.

[10] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," Int. J. Computer Vision, Vol. 8, No. 2, pp. 99-111, 1992.

[11] P. Sinha, "Object recognition via image invariants: a case study. Investigative ophthalmology and visual science," Vol. 35, pp. 1735-1740, 1994.

[12] S. McKenna and S. Gong, "Tracking faces," Proc. the 2nd Int. Conf. on Automatic Face and Gesture Recognition, Killington, Vermont, 1996.

[13] G. Wyszecki and W.S. Styles, "Color Science: Concepts and Methods, Quantitative Data and Formulae," Second Edition, John Wiley & Sons, New York, 1982.

[14] R.M. Haralick and G.L. Kelly, "Pattern recognition with measurement space and spatial clustering for multiple images," Proceedings of IEEE, Vol. 57, No. 4, pp. 654-665, 1969.

[15] Y. Ohta, T. Kanade and T. Sakai, "Color information for region segmentation," *Computer Graphics and Image Processing*, Vol. 13, No. 3, pp.222-241, July 1980.

[16] M.J. Swain and D.H. Ballard, "Color indexing," International Journal of Computer Vision. Vol. 7, No.1, pp. 11-32, 1991.

[17] B.V. Funt and G.D. Finlayson, "Color constant color indexing," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, No. 5, pp. 522-529, 1995

[18] J.D. Crisman and C.E. Thorpe, "SCARF: a color vision system that tracks roads and intersections," IEEE Trans. Robot. Autom., Vol. 9, No. 1, pp. 49-58, 1993.

[19] S.A. Shafer, "Optical phenomena in computer vision," Prec. Canadian Soc. Computational Studies of Intelligence, pp. 572-577, 1984.

[20] G.J. Klinker, S.A. Shafer, and T. Kanade, "Using a color reflection model to separate highlights from object color," Proc. ICCV, pp. 145-150, 1987.

[21] R. Gershon, "The Use of Color in Computational Vision," Ph.D thesis, University of Toronto, 1987.

[22] D.H. Brainard, B.A. Wandell, and E.-J. Chichilnisky, "Color constancy: from physics to appearance," Current Directions in Psychological Science, Vol. 2, No. 5, pp. 165-170, 1993.

[23] D. Forsyth, "A novel algorithm for color constancy," International Journal of Computer Vision. Vol. 5, No. 1, pp.5-36, 1990.

[24] W. J. Krazanowski, "Principles of Multivariate Analysis," Clarendon Press, Oxford, 1988.

[25] J. D. Jobson, "Applied Multivariate Data Analysis," Vol. I: Reregression and Experimental Design, Springer-Verlag, 1991.

[26] J. N. Srivastava, "On testing hypertheses regarding a class of covariance structures," Psychometrika, Vol. 31, No. 2, pp. 147-164, 1966.

[27] T.W. Anderson, "Statistical inference for covariance matrices with linear structure," Proceedings of the Second International Symposium on Multivariate Analysis, pp. 55-66, 1969.

[28] T.W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure," The Annals of Statistics, Vol. 1, No. 1, 135-141, 1973.

[29] G.S. Rogers and D.L. Young, "Explicit maximum likelihood for certain patterned covariance matrices," Communications in Statistics (Part A), Vol. 6, No. 2, pp121-133, 1977.

[30] D.K Dey and A.E. Gelfand, "Improved estimation of a patterned covariance matrix," Journal of Multivariate Analysis, Vol. 31, pp. 107-116, 1989.

[31] J. Yang, L. Wu, and A. Waibel, "Focus of attention: towards low bitrate video teleconferencing," Proceedings of 1996 IEEE International Conference on Image Processing (Lausanne, Switzerland), Vol. 2, pp. 97-100.

[32] R. Stiefelhagen, J. Yang, and A. Waibel, "A model-based gaze tracking system," To appear on International Journal on Artificial Intelligence Tools.

[33] R. Stiefelhagen, U. Meier, and J. Yang, "Real-time Lip-tracking for Lipreading," Proceedings of Eurospeech'97.