Skin-Color Modeling and Adaptation

Jie Yang, Weier Lu, Alex Waibel

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA

Abstract. This paper studies a statistical skin-color model and its adaptation. It is revealed that (1) human skin colors cluster in a small region in a color space; (2) the variance of a skin color cluster can be reduced by intensity normalization, and (3) 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 tracking human faces under different lighting conditions. The parameters of the model are adapted based on the maximum likelihood criterion. The model has been successfully applied to a real-time face tracker and other applications.

1 Introduction

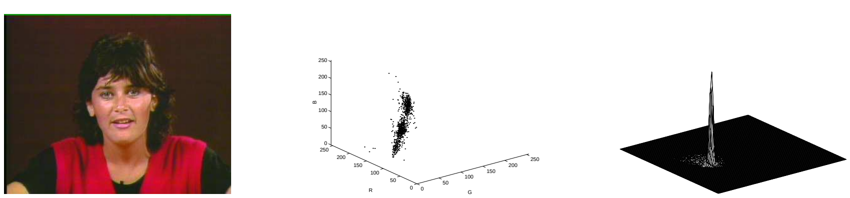
Human face perception is currently an active research area in the computer vision community. Locating and tracking human faces is a prerequisite for face recognition and/or facial expressions analysis, although it is often assumed that a normalized face image is available. Facial features, such as eyes, 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.

Color is another feature on human faces. Much research has been directed to understanding and making use of color information. Color has been long used for recognition and segmentation [1, 2] and recently has been successfully used face locating and tracking [3, 4, 5]. 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 [6]. Using color as a feature for tracking human faces 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.

In this paper, we quantitatively investigate human skin color distributions. 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.

2 Skin-Color Distributions

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 [2]. 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 [7]. It has since been shown that the colors do not fall randomly in a plane, but form clusters at specific points [8]. It has been further observed that (1) human skin colors cluster in a small region in a color space; (2) human skin colors differ more in intensity than in colors, and (3) under a certain lighting condition, a skin-color distribution can be characterized by a multivariate normal distribution in the normalized color space [5]. The Figure 1 shows a face image, the skin-color occurrences in the RGB color space (256x256x256), and the skin color distribution in the normalized color space. In the following section, we justify these observations by quantitative analysis goodness-of-fit techniques.



(a) Face image (color!) (b) Skin-color occurrences (c) Distribution

Fig. 1. An example of a human skin-color cluster and distribution

3 Quantitative Analysis and Goodness-of-Fit Test

We have built up a database which contains about 1000 face images down-loaded from the Internet and taken from our laboratory. This database covers face images of people in different races (Caucasian, African American, and Asian), genders, and the lighting conditions.

3.1 Data Analysis

Figure 2 shows in the RGB space skin color occurrences of 48 human faces randomly selected from the database. 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.

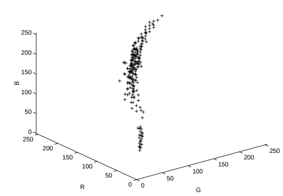


Fig. 2. Skin color cluster of 48 faces (mean values: $m_R = 188.9069$, $m_G = 142.9157$, $m_B = 115.1863$, and variances: $\sigma_R = 58.3542$, $\sigma_G = 45.3306$, $\sigma_B = 43.397$).

We further studied the effect of illumination on skin color clusters by comparing variances of skin color clusters in RGB space and the normalized color space. It has been found that variances of skin color clusters can be greatly reduced by the color normalization. Table 1 shows mean values and variances of the same color cluster in different color spaces.

3.2 Goodness-of-Fit Tests

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 have tested whether the measured data of a sample do indeed have a normal distribution by goodness-of-fit techniques. Goodness-of-fit tests examine the conformity of the observed data's empirical

Table 1. Comparison of mean and variance

	RGB Space	Normalized Color Space
Mean	$m_R = 234.29$ $m_G = 185.72$ $m_B = 151.11$	$m_r = 104.22$ $m_g = 81.59$
Variance	$\sigma_R = 26.77$ $\sigma_G = 30.41$ $\sigma_B = 25.68$	$\sigma_r = 4.93$ $\sigma_g = 3.89$

distribution function with a posited theoretical distribution function. The methods of performing a test can be an analytic or a graphic approach. In the graphic approach, the most common method is *Quantile-Quantile* plot (or Q-Q plot). We will use this method to test skin-color distributions. To perform a goodness-of-fit test, we first need to formulate a null hypothesis.

NULL hypothesis:
human skin-color is normally distributed in a normalized bivariate space.

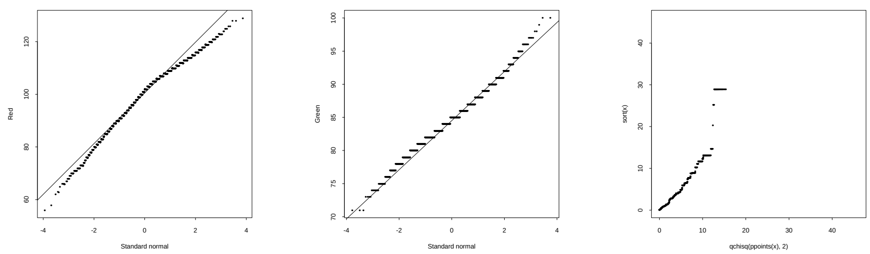
An immediate difficulty of the task is that there is no commonly agreed analytical tool available to test the normality of a bivariate distribution [9]. In order to solve the problem, we used a two step strategy: first we test the marginal distribution and then test multivariate distribution. If a multivariate distribution is a normal distribution, its marginal distributions must be normal distributions. If the marginal distributions fail to pass the normality test, there is no need to test the multivariate distribution.

The basic idea of the Q-Q plot is to use the cumulative probability of the sampling data against that of the tested distribution. A straight line indicates that we cannot reject the null hypothesis (the interested readers are referred to textbooks on the subject, e.g. [9]). When we do marginal test, we test each variable separately against the normal distribution. When we test the bivariate distribution, we test the transformed variable against χ^2 distribution [9].

We first tested marginal distributions. As the results, we could not reject the null hypothesis. Figure 3(a) and (b) are the Q-Q plots of an African American's marginal distributions. Both the normalized "r" and "g" are straight lines. We then tested bivariate distributions. Again, we cannot reject the null hypothesis. Figure 3(c) is an example of the Q-Q plot for a Caucasian skin-color distribution. The plot is a straight line except a few outliers. Therefore, we have verified that, under a certain lighting condition, the skin-color distribution of an individual can be characterized by a Gaussian distribution.

4 Maximum Likelihood Adaptation

Although under a certain lighting condition, 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. There are



(a) Marginal test (r) of an African American (b) Marginal test (g) of an African American (c) χ^2 test of a Caucasian

Fig. 3. Examples of Q-Q plots

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. Adaptive approach, on the other hand, is to transform the previous developed color model into the new environment. Since the Gaussian model has only a few parameters, it is possible to update them in real-time. Because the linear combination of Gaussian distributions is still a Gaussian distribution, we can consider the current Gaussian distribution is a combination of the previous distributions. One way of adaptation is to use the linear combination of known parameters to predict, or, approximate the new parameters, i.e.,

$$\hat{\mu} = \sum_{k=1}^{r} \alpha_k \mathbf{m_k}, \quad \hat{\Sigma} = \sum_{k=1}^{r} \beta_k S_k, \tag{1}$$

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

We will use the maximum likelihood criterion to find the best set of coefficients for the prediction.

Let the sample mean and variance be

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x_k}, \quad C = \frac{1}{N} \sum_{k=1}^{N} (\mathbf{x_k} - \bar{\mathbf{x}})(\mathbf{x_k} - \bar{\mathbf{x}})' \tag{2}$$

The logarithm of the likelihood function can be written as:

$$\log L = -N \log(2\pi) - \frac{1}{2} N \log \left| \Sigma^{-1} \right| - \frac{1}{2} N tr \; \Sigma^{-1} C - \frac{1}{2} N (\bar{\mathbf{x}} - \mu)' \Sigma^{-\mathbf{1}} (\bar{\mathbf{x}} - \mu). \tag{3}$$

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. 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. By setting the derivatives of the likelihood function (3) with respect to $\alpha_k, k = 1, \ldots, r$, to 0, the equations for the maximum likelihood estimates are

$$\sum_{k=1}^{r} \mathbf{m_j}' \Sigma^{-1} \mathbf{m_k} \hat{\alpha_k} = \mathbf{m_j}' \Sigma^{-1} \bar{\mathbf{x}}, \quad j = 1, \ldots, r \tag{4}$$

We can obtain α_k by solving the equation (4).

4.2 Mean Vector and Covariance Matrix Adaptation

In this case, both mean vector and covariance matrix are assumed to be a linear combination of the previous parameters. Many researchers in the field of statistics have studied this problem. In general, explicit solutions for this problem do not exist and estimates must be performed by iterative numerical techniques.

In fact, because the two sets of estimates are asymptotically independent, each set of parameters can be estimated as when the other set of parameters is known. In the following we present an EM algorithm based on the estimate procedure proposed by Anderson [10]. 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.

Algorithm

1. Initialization

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

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

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

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

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

2. Iteration

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

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

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

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

$$j = 1, \ldots, r,$$

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

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 [10].

4.3 Applications

Th adaptive skin-color model has been applied to many applications. The model plays a key role in the real-time face tracker [5]. 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 been applied to other applications such as tele-conferencing [11], gaze tracking [12], and lip-reading [13].

5 Conclusions

We have proposed a statistical skin-color model for tracking human faces in real-time. We have shown that the variance of skin color clusters 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.

Acknowledgements

This research was sponsored by the Advanced Research Projects Agency under the Department of the Navy, Naval Research Office under grant number N00014-93-1-0806.

References

1. 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.
2. M.J. Swain and D.H. Ballard, "Color indexing," International Journal of Computer Vision. Vol. 7, No.1, pp. 11-32, 1991.
3. 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.
4. 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.
5. 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 ("Tracking human faces in real-time," Technical Report CMU-CS-95-210, CS department, CMU, 1995).
6. G. Wyszecki and W.S. Styles, "Color Science: Concepts and Methods, Quantitative Data and Formulae," Second Edition, John Wiley & Sons, New York, 1982.
7. S.A. Shafer, "Optical phenomena in computer vision," Prec. Canadian Soc. Computational Studies of Intelligence, pp. 572-577, 1984.
8. 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.
9. J. D. Jobson, "Applied Multivariate Data Analysis," Vol. I: Reregression and Experimental Design, Springer-Verlag, 1991.
10. T.W. Anderson, "Asymptotically efficient estimation of covariance matrices with linear structure," The Annals of Statistics, Vol. 1, No. 1, 135-141, 1973.
11. J. Yang, L. Wu, and A. Waibel, "Focus of attention: towards low bitrate video tele-conferencing," Proceedings of 1996 IEEE International Conference on Image Processing (Lausanne, Switzerland), Vol. 2, pp. 97-100.
12. R. Stiefelhagen, J. Yang, and A. Waibel, "A model-based gaze tracking system," To appear on International Journal on Artificial Intelligence Tools.
13. U. Meier, R. Stiefelhagen, J. Yang, "Preprocessing of Visual Speech under Real World Conditions" European Tutorial & Research Workshop on Audio-Visual Speech Processing: Computational & Cognitive Science Approaches (AVSP 97).