AUTOMATIC SKIN PIXEL SELECTION AND SKIN COLOR CLASSIFICATION

Sangho Yoon
Department of Electrical Engineering
Stanford University, CA 94305

Michael Harville, Harlyn Baker, Nina Bhatii
Hewlett-Packard Laboratories
Palo Alto, CA 94304

ABSTRACT

We describe an automatic, low-cost method for classifying skin color, independent of lighting and imaging device characteristics, using consumer digital cameras and a simple color calibration target. After color normalization and face detection are performed as described in prior work, pixels within the face region are clustered in color space in an unsupervised fashion, and a Gaussian mixture model (GMM) of the person's skin color is formed from the pixels belonging to the densest clusters containing at least some minimum fraction of the total pixels. This technique allows accurate modeling of non-uniformities in skin tone that are common in individuals, while avoiding contamination from shadows, specularities, eyes, lips, hair, and background. We incorporate these models into a skin color classification framework with improved performance over prior work. Given a set of exemplar face images with skin color labels assigned by an expert, we predict the label that would be chosen by the same expert for a new, test image by comparison of the respective GMMs of the test image and each exemplar. Specifically, we select the label of the exemplar image whose GMM has smallest Kullback-Leibler divergence from that of the test image.

1. INTRODUCTION

Most prior image processing work on skin color classification either does not discount illuminant or camera properties, requires expensive equipment, or involves a tedious process with trained operators [11, 12, 13]. Recently, Harville et al. [8] presented fast techniques for measuring and classifying facial skin color from a single, casually posed digital camera image. In their system, color calibration and correction based upon a handheld color pattern (see Figure 1) was applied to acquired images to account for the effects of the scene illuminant and camera system. Skin pixels were automatically selected by simple luminance filtering within a subregion of the image window returned by a standard face detector. The skin color of a person was represented as the mean of these pixels, and a single Gaussian was used to model the distribution of these means for all subjects having the same skin color label. Skin color classification of a test subject was performed by selecting the label of the Gaussian with smallest Mahalanobis distance to the mean color of the subject. This technique facilitated applications in several areas that would benefit from objective measurement of human skin tone:

- Medicine: for quantification of skin erythema, lesions, ultraviolet radiation effects, and other phenomena
- Computer graphics: for more accurate rendering of people in video-conferencing, or for altering their appearance
- Fashion: for automated suggestion of personal appearance products, such as clothing, that complement skin tone
- Biometrics: to aid in person recognition within small groups, or other systems in which skin color determination is useful

The color sampling and classification algorithms in [8] were simple and showed good results for many labelings of skin color. In this paper, we focus on improving selection of representative skin pixels, and on enabling skin color classification to handle more complex skin color descriptions. We follow the same color calibration and correction techniques as in [8], but we select skin pixels through analysis of probability density functions (PDFs) of pixels in each face image. More precisely, a Gaussian mixture model (GMM) is fit to the face pixel colors, and some components are then eliminated as unlikely to be representative of skin. For classification using these more complex skin color representations, we use the Kullback-Leibler (KL) divergence [2] to measure similarities between GMMs. The skin color of a query face image is assigned the class label of the closest (in terms of the KL divergence) image in the training set.

The rest of paper is organized as follows. In Section 2, we discuss color correction, face detection, and pixel color clustering algorithms. In Section 3, we present our skin pixel selection and color classification algorithm. We show experimental results in Section 4, and conclude in Section 5.

2. BACKGROUND

2.1. Color Correction and Skin Region Selection

We would like our methods to be applicable to digital images captured by most cameras, under a wide range of illuminants. Since it is difficult, expensive, or time-consuming for end users to control illumination (particularly for outdoor mobile contexts) or pre-calibrate their camera, we instead employ the color correction method of [8], which requires the presence in the image of a detectable pattern of known colors. For some applications, this pattern might take the form of a paper color chart distributed to end users, who capture an image of themselves with their own camera.

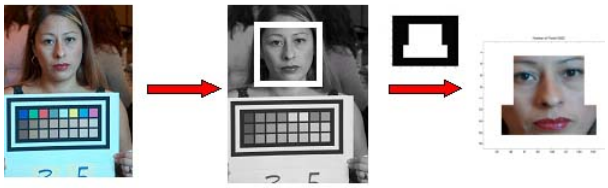


Fig. 1. Color calibration and correction, automatic face detection, and masking

Figure 1 shows an image of a user holding an example color chart. It contains 3 rows of 8 color patches set against a black background, wrapped by a white then a black frame. The top row contains primary and secondary colors for a general scene tone balancing, and two shades of gray for white balance. The other two rows contain 16 patches representative of the range of human skin color. The color calibration chart is detected by analyzing zero crossings of the Laplacian of a smoothed luminance version of the image. The image color within each patch is sampled, and compared to the known reference values to compute a 4x3 linear transform for remapping colors to a reference space. Due to the many skin tones on the color chart, this remapping is more accurate for skin tones (our color range of interest) than other colors.

Once colors are corrected, we apply the Viola-Jones face detector [15] to obtain a square region bounding the face in the image. When multiple faces are detected, the largest is used. A mask is then applied within the square to further reduce the area of our interest, better excluding background and hair. Figure 1 shows an example of this.

2.2. Pixel Color Clustering and KL Divergence

We use vector quantization (VQ) [5] to cluster the colors of pixels within the selected face region. VQ is a clustering algorithm because it represents an input vector by one of a predetermined set of patterns (or codewords) on the basis of which pattern is closest to the given input vector. The encoder and decoder in VQ are associated with partitions (clusters) and codewords (cluster centers), respectively. VQ can also be viewed as fitting a model, where partition cells are represented by their conditional probability density functions and where prior probabilities are weights.

In particular, we are interested in fitting Gaussian mixture models to data within a VQ framework [1]. In this approach, the PDF of an input vector X is represented as a weighted collection of Gaussians:

$$f(X) = \sum_{k=1}^{N} p_k f(X|k) \quad (1)$$

where N, p_k, and $f_k(X)$ are the number of clusters, the prior probability of cluster k, and the conditional PDF of cluster k, respectively. The conditional PDF $f_k(X)$ is expressed as

$$f(X|k) = \frac{\exp\left(-\frac{1}{2}(X-m_k)^t \Sigma_k^{-1}(X-m_k)\right)}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} \quad (2)$$

where $X \in \mathcal{R}^p$, and m_k and Σ_k are the mean and the covariance matrix of cluster k, respectively.

GMMs have been used in various areas in signal processing and shown to be robust [7]. The EM algorithm [3] is the most popular approach to fitting a GMM to data, but we instead use the alternative Lloyd algorithm [10][1]. The main difference between them is that the EM algorithm makes soft decisions for input data, whereas the Lloyd algorithm makes hard decisions. The Lloyd algorithm first assigns data to the closest cluster centers, next updates cluster centers and then iterate these two steps until convergence is reached.

Once we have GMMs for the colors of different face images, we can measure similarity between them using the KL divergence. The KL divergence is a general means of measuring similarity between two PDFs, and has been used successfully in image retrieval [14][6]. For two N-component GMMs $p(x) = \sum_{i=1}^{N} p_i f_p(x|i)$ and $q(x) = \sum_{j=1}^{N} q_j f_q(x|j)$, the KL divergence may be expressed as:

$$\begin{aligned} D(p(x)||q(x)) &= \int p(x)\log\frac{p(x)}{q(x)}dx \quad (3) \\ &= \int \sum_{i=1}^{N} p_i f_p(x|i)\log\frac{\sum_{i=1}^{N} p_i f_p(x|i)}{x\sum_{j=1}^{N} q_i f_q(x|j)}dx \end{aligned}$$

Since there is no closed form solution to Eq. (3), we use an approximation in [6, 14]:

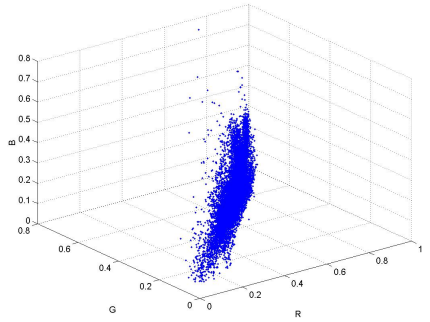
$$\begin{aligned} D(p(x)||q(x)) &\simeq \sum_{i=1}^{N} p_i \text{min}_j \left(D(f_p(x|i)||f_q(x|j)) + \log(\frac{p_i}{q_j})\right) \\ &= \sum_{i=1}^{N} p_i \left(D(f_p(x|i)||f_q(x|\pi(i))) + \log(\frac{p_i}{q_{\pi(i)}})\right) \end{aligned}$$

where $\pi(i) = \text{argmin}_j \left(D(f_p(x|i)||f_q(x|j)) + log(\frac{p_i}{q_j})\right)$. Since $f_p(x|i)$ and $f_q(x|j)$ are multivariate Gaussians, there is a closed form solution to the KL divergence $D(f_q(x|j)||f_q(x|j))$, and we can compute the approximation above using means and covariance matrices. See [2, 6] for more details.

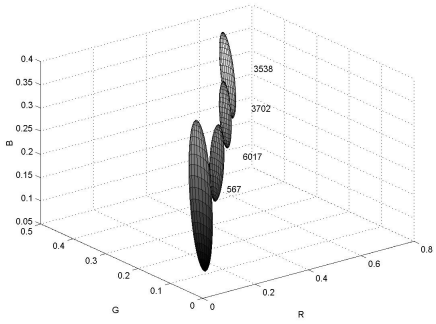
3. MODELING AND CLASSIFYING SKIN COLOR

Even after applying a mask to the detected face as shown in Figure 1, the selected region still contains hair, lips, eyes, and background. We wish to avoid these while selecting pixels representative of skin color. Further, the distribution of skin pixel colors for a single person is often complex, due to variations in pigmentation, sun damage, and other phenomena. Within a given image of a person, the distribution is further complicated by specularities and shadows.

We therefore apply the clustering algorithm of Section 2.2 to the pixel colors inside the masked area, both to help identify distractors such as hair and background, as well as to allow a detailed representation of a person's skin tone. Clustering not only gives us a partition of the pixels inside the mask,



(a) Raw distribution



(b) Partitioned into four clusters: numbers represent densities

Fig. 2. Distribution and Clustering

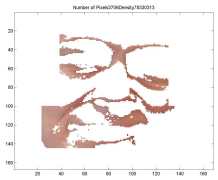
but also a GMM fitted to the pixels of the inside the mask. Each cluster is represented by a multivariate Gaussian, so that a weighted collection of them approximates the distribution of the pixels inside the mask. While some of these clusters correspond to non-skin distractors, or skin under significant specularity or shadow, the remainder are used as an estimate of the underlying PDF of the person's skin color.

Figure 2(a) shows the original distribution of pixels inside the masked face area of a single subject image, and Figure 2(b) shows the color clustering result. Note that most pixels inside the masked area in Figure 1 are skin pixels. This was found to be true across all images we used. Based on this observation, we can assume that non-skin pixels comprise a small portion of those pixels inside the mask. In this paper, we use a fixed number (N=4) of clusters for each image.

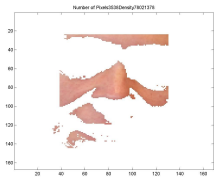
Because the number of non-skin pixels inside the masked area is much smaller than skin pixels, and because non-skin pixels tend to be more widely spread in color space than skin pixels, we expect the set of skin pixels to be more densely clustered than the set of non-skin pixels. We therefore refine our selection of skin pixels by choosing clusters according to a cluster density criterion. After clusters are produced as in Section 2.2, we order them by density:

$$\text{Dens(k)} \quad = \quad \frac{n_k}{|\Sigma_k|} \tag{4}$$

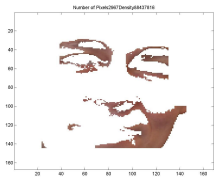
where n_k is the number of pixels assigned to cluster $k(1 \leq k \leq N)$, and $|\Sigma_k|$ is the determinant of Σ_k. Among all



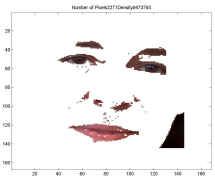
(a) 1^{st} densest cluster image



(b) 2^{nd} densest cluster image



(c) 3^{rd} densest cluster image



(d) 4^{th} densest cluster image

Fig. 3. Cluster member pixels, in order of density

clusters containing at least a small minimum percentage of pixels inside the masked area, we choose those with highest density as being representative of the person's skin color. The requirement of a minimum percentage prevents selection of very small clusters representing outlier colors in the masked area (e.g. a blue earring). We experimented with the number of densest clusters selected, as described in Section 3.

Each face image is represented by a weighted collection of the Gaussians corresponding to the selected skin pixel clusters. To classify skin color for a query image, we use the KL divergence to measure similarities between the GMM skin color representations of the query face image and the face images in our database. The query image is assigned the class label of the database image with smallest KL divergence.

4. EXPERIMENTAL RESULTS

To evaluate our methods, we used a data set of 142 images acquired by multiple cameras under various lighting conditions. The variation of lighting and cameras was greater for this data set than in [8]. Each image was assigned a ground truth skin color class label according to which of the 16 skin-toned patches in bottom two rows of the color calibration target of Figure 1 most closely matched the mean color of representative skin pixels selected in the image by a human expert. To evaluate skin color classification performance, we use the leave-one-out cross validation method [9] with two performance measures: average correct rank and ROC curve [4]. In this technique, we treat each image in turn as a query image, using the other 141 images as the image database. We order the database images according to similarity using the KL divergence, and identify the rank of the correct class label on this list. The rank of the correct class is averaged across all 142 trials to obtain our average correct rank metric, while an ROC curve shows statistics on how often the correct label was in the top two choices, top three choices, and so on.

We partitioned each image into four clusters, and experimented with varying the number of densest clusters selected to represent the person's skin color. Specifically, we tried using only the densest cluster, the two most dense clusters, the top three, and all four. For our data set, the best skin color classification performance was obtained by using the three densest clusters for each image to represent skin color. This indicates that the least dense cluster almost always contained non-skin, whereas the second and third most dense clusters too often contained important skin color information to be omitted from the model. Figure 3 shows the pixels belonging to four clusters of a single example image, and indicates that only the 4^{th} cluster contains non-skin pixels.

We compare our classification performance with that of method [8]. Figure 4 shows ROC curves and Table 1 shows average correct rank results, both of which indicate improved performance. We believe the performance improvement would be even more significant if we were to use skin color class labels whose basis is more complex than proximity to skin color patches on the calibration target. We plan to repeat our experiments for image data assigned color class labels by an expert who accounts for fine variation and mottling in skin tone, and we believe the greater flexibility of the skin color representations presented here will be critical for good performance.

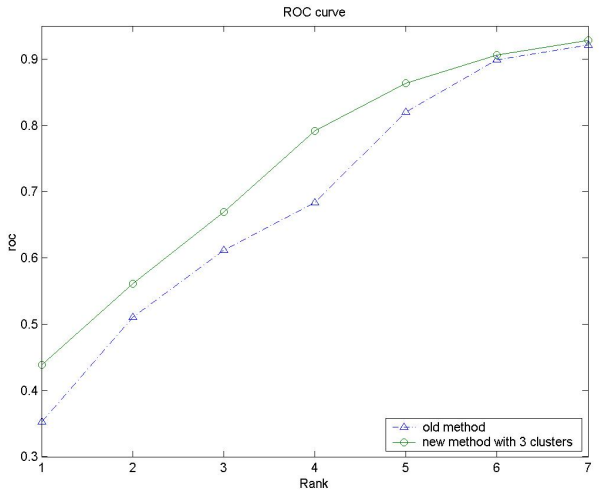


Fig. 4. ROC performance comparison with method [8]

Table 1. Average correct rank

	old method	new method
Average correct rank	3.2158	2.9784

5. CONCLUSIONS

We believe we have improved upon methods for objective skin color measurement and classification, through use of GMMs that estimate the PDF of skin colors for individuals. We show that analysis of cluster density in color space is useful in forming these models. Skin color classification on a challenging data set was shown to be improved over prior techniques that represent skin color using only the mean. For future work, we are investigating more sophisticated methods to choose clusters instead of choosing some fixed number of clusters.

6. ACKNOWLEDGEMENT

We thank Prof. Robert M. Gray from Stanford and Prof. Sabine Susstrunk from EPFL for their valuable comments.

7. REFERENCES

[1] A. K. Aiyer et al. "Lloyd Clustering of Gauss Mixture Models for Image Compression and Classification", *Signal Processing: Image Communication*, Vol. 20(5), pp. 459-485, 2005.

[2] T. Cover and J. Thomas, Elements of Information Theory, Wiley Series in Telecommunications, *John Wiley and Sons*, 1991.

[3] Dempster et al., "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *J. of the Royal Statis. Soc.*, Series B, Vol. 39, 1977, p. 1-38.

[4] R. Duda, P. Hart, and D. Stork, "Pattern Classification," 2nd ed., *John Wiley and Sons*, 2001.

[5] A. Gersho and R. Gray, "Vector Quantization and Signal Compression," *Kluwer Academic Press*, 1992.

[6] J. Goldberger et al. "An efficient image similarity measure based on approximations of KL-divergence between two gaussian mixtures," *Proc. ICCV*, 2003.

[7] R.M. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization," *IEEE Trans. Inform. Theory, Vol. 49*, pp. 1204-1217, May, 2003.

[8] M. Harville et al.,"Consistent Image-Based Measurement and Classification of Skin Color," *Proc. ICIP*, 2001.

[9] T. Hastie et al.," The Elements of statistical learning," *Springer-Verlag*, 2001

[10] S. Lloyd, "Least square quantization in PCM," *IEEE Transactions on Information Theory*, IT-28(2):129-137, March 1982.

[11] M. Nischik, C. Forster, "Analysis of skin erythema using true-color images," *IEEE Trans. Medical Imaging(16)*, no. 6, 1997.

[12] H. Takiwaki, "Measurement of skin color: practical application and theoretical considerations," *J. Med. Invest(44)*, 1998.

[13] Y. Vander Haeghen, J. Naeyeart, I. Lemahieu, "Consistent digital color image acquisition of the skin," *Intl. Conf. Eng. in Med. and Bio.*, 1998.

[14] N. Vasconcelos, "On the complexity of probablistic Image retrieval," *Proc. ICCV*, 1999.

[15] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. CVPR*, 2001.