A Person and Context Specific Approach for Skin Color Classification

Matthias Wimmer
wimmerm@cs.tum.edu

Bernd Radig
radig@cs.tum.edu

Michael Beetz
beetz@cs.tum.edu

Technische Universität München, Garching bei München, Germany

Abstract

Skin color is an important feature of faces. Various applications benefit from robust skin color detection. Depending on camera settings, illumination, shadows, people's tans, and ethnic groups skin color looks differently, which is a challenging aspect for detecting it automatically.

In this paper, we present an approach that uses a high level vision module to detect an image specific skin color model. This model is then used to adapt parametric skin color classifiers to the processed image. This approach is capable to distinguish skin color from extremely similar colors, such as lip color or eyebrow color. Its high speed and high accuracy make it appropriate for real time applications such as face tracking and recognition of facial expressions.
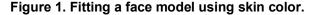
1. Introduction

Robust skin color detection supports computer vision applications such as face detection, mimic recognition, person identification, hand gesture detection, and image content filtering [2,3,8]. Applications that rely on exactly fitting a contour model to faces, such as mimic recognition, benefit from correctly detected skin color regions, see Figure 1.

Classifying skin color is challenging, because it occupies a large cluster within color space. The reason is that camera types and settings, illumination conditions, as well as people's tans and ethnic groups make skin color vary significantly. However, within one image skin color looks similarly, because most of the above conditions are fixed. An image specific skin color model describes those conditions. It is acquired with the a priori knowledge of some skin color pixels from the image. Using this model, parametric skin color classifiers are adapted to the processed image, in order to improve their classification accuracy.

Previous work focuses on detecting a skin color model via low level techniques such as color

segmentation, background subtraction, or histogram prediction [4]. We use the commonly known and widely accepted face detector of Viola and Jones [7] and extend it with an empirically obtained skin color mask for this task. Since the face detector works with gray value images it is not influenced by the color distribution. We consider this approach as a combination of a high level vision module with a low level vision module.

Figure 1. Fitting a face model using skin color.

Our approach is divided into two phases. First, we acquire the image specific skin color model. This is done by locating the face, applying the skin color mask, extracting a small number of skin color pixels, and assembling the skin color model. Second, we adapt a parametric skin color classifier to the skin color model.

The remainder of the paper is organized as follows: The next section gives a short overview of related work. Section 3 explains the image specific skin color model and how we obtain it. Section 4 introduces two parametric skin color classifiers and their adaptation via the skin color model, and Section 5 demonstrates experimental results. We conclude with, and discuss future work in Section 6.

2. Related work

Broadhurst and Baker [1] introduce a framework where high level modules feed their knowledge back to parametric low level modules in order to improve the performance of the overall system. They evaluate their approach in several applications, such as color based face tracking or edge based lane tracking.

Vezhnevets et al. [6] give a comprehensive overview about the work within the area of skin color detection that has been done during the last decade. They describe the main color spaces and the skin color models and categorize the classification techniques. Our approach contributes to the category *dynamic skin distribution models*, which describes parametric skin color classifiers that are adapted to the image conditions prior to the classification step. Soriano et al. [5] define their skin color model as a cluster with a distinct shape in color space. This model is adapted to the illumination conditions by evaluating the image's histogram. Sigal et al. [4] predict their skin color histogram using a second order Markov model based on the current segmentation.

Viola and Jones [7] create a face detector that works with gray value images. Their detector uses a boosted cascade of classifiers that evaluate rectangular features and sum up the gray values inside. It basically makes use of the brighter and darker parts within faces and delivers squares around the detected faces.

3. The image specific skin color model

Skin color can not be modeled properly in the RGB color space, because the corresponding skin color cluster is very large and incompact. A common way to cope with that is to use the normalized RGB color space (*NRGB*), which uses the proportional part of each color, see (1). *b* is omitted because it can be calculated from the other components [6]. The color of a pixel p is represented by the color vector $c_p = (r, g, base)^T$.

$$
\begin{aligned}
base &= R + G + B, & r &= R / base \\
g &= G / base, & b &= B / base
\end{aligned}
\quad (1)
$$

The image specific skin color model is represented by the mean μ and the covariance matrix S of the color vectors of all skin color pixels, see (2). I is the number of skin color pixels, var stands for the variance and cov stands for the covariance.

$$
\mu = \begin{pmatrix} \mu_r \\ \mu_g \\ \mu_{base} \end{pmatrix} = \frac{1}{I} \sum_{i=1}^{I} c_i
$$

$$
S = \begin{pmatrix} var_r & cov_{r,g} & cov_{r,base} \\ cov_{r,g} & var_g & cov_{g,base} \\ cov_{r,base} & cov_{g,base} & var_{base} \end{pmatrix} = \frac{1}{I-1} \sum_{i=1}^{I} (c_i - \mu)(c_i - \mu)^T
\quad (2)
$$

We approximate the image specific skin color model in the following way: First, the face detector delivers a square around the detected face, called ROI. Then the skin color mask is applied to that ROI which extracts a small number of skin color pixels. The image

specific skin color model is set up using only those pixels.

3.1. Skin color masks

A skin color mask enables a face detector to extract skin color pixels from the ROI. It is a two-dimensional matrix that specifies the probability of skin color for each pixel within the ROI after being scaled to its size. Runtime performance is increased by only taking those probability entries into account that exceed a given threshold value. A skin color mask exists for any face detector; however, we show its benefit for the Viola and Jones face detector.

A skin color mask M is learned via training images whose skin color pixels are known. Therefore, each pixel needs to be manually labeled as *skin* or *non-skin*. We use a set of K training images that show faces of different persons out of different ethnic groups, different illumination conditions, different sizes, and slightly different poses. Those training images originate from well known face detectors, the Boston University skin color database [4], and various web pages.
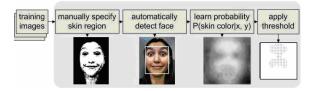

Figure 2. Acquiring a skin color mask.

M is an $n_1 \times n_2$ matrix whose entries are called $m_{i,j} \in [0..1]$. In this work we take $n_1=n_2=24$ as a reasonable compromise between accuracy and runtime performance. We apply the face detector to each image k and receive a region of interest roi_k which is divided into $n_1 \times n_2$ cells $f_{k,i,j}$ with $1 \leq i \leq n_1$ and $1 \leq j \leq n_2$. The likelihood for skin color within cell $f_{k,i,j}$ is expressed by $s_{k,i,j}$. Finally, we calculate the entries of the skin color mask $m_{i,j}$, see (3). The entire procedure is depicted in Figure 2.

$$
s_{k,i,j} = \frac{number\ of\ skin\ color\ pixels\ in\ f_{k,i,j}}{total\ number\ of\ pixels\ in\ f_{k,i,j}}
$$

$$
m_{i,j} = \frac{1}{K} \sum_{k=0}^{K-1} s_{k,i,j}
\quad (3)
$$

3.2. Empirical results

We base the evaluation on the Boston University skin color database [4]. It consists of 21 image

sequences that are taken from Hollywood movies. Their length varies between 49 and 349 frames and they show persons in natural activities such as talking, walking, or working. They include various illumination conditions and people from various ethnic groups. Each pixel is annotated with *skin*, *non-skin*, or *don't care*. Since our approach works with a face locator for frontal faces we use only those video sequences that include frontal faces. Table 1 depicts the accuracy of our approach compared to a low level method. Both obtain μ, and evaluate the relative distance to the ground truth of μ. The ground truth is obtained by evaluating all pixels that are marked as *skin*. Obviously our approach approximates the ground truth up to seven times closer.

Table 1: Two approaches obtain *μ*.

seq	our approach			color segmentation		
	r	g	base	r	g	base
2	0.8%	0.4%	6.5%	0.9%	0.8%	11.1%
4	1.0%	0.4%	0.5%	7.1%	5.7%	4.6%
6	0.1%	0.0%	2.8%	5.7%	1.4%	16.9%
7	1.0%	0.6%	3.9%	5.3%	5.8%	15.2%
8	1.1%	0.1%	7.3%	5.1%	1.2%	5.5%
9	0.7%	0.4%	1.2%	0.9%	2.2%	10.5%
10	0.4%	0.3%	3.2%	6.0%	1.2%	16.5%
11	0.4%	0.9%	4.1%	3.6%	1.1%	1.5%
15	0.2%	0.1%	3.1%	4.1%	1.3%	1.4%
16	0.5%	0.1%	5.0%	3.3%	1.8%	25.8%
18	0.8%	0.5%	8.1%	4.8%	1.6%	7.3%
avg	0.6%	0.3%	4.2%	4.3%	2.2%	10.6%

4. Parametric skin color classifiers

The image specific skin color model adapts parametric skin color classifiers to the processed image. During the last two decades various skin color classifiers have been proposed. Some of them are parametric and the others can be made parametric with ease. We choose two simple classifiers and show the benefit of our work.

Cuboid cluster: Skin color is defined to be located in a cuboidal cluster in NRGB. The lower and upper bounds of the cuboid (l_r, l_g, l_{base}, u_r, u_g, u_{base}) are calculated from the parameters $\mu\psi$ and $S\psi$ of the image specific skin color model, see formulae below. That approach adapts this classifier to the image conditions. The standard deviation $\sigma_i = \sqrt{\text{var}_i}$ is extracted from the covariance matrix S.

$$
\begin{aligned}
l_r &= \mu_r - 2\,\sigma_r, & u_r &= \mu_r + 2\,\sigma_r \\
l_g &= \mu_g - 2\,\sigma_g, & u_g &= \mu_g + 2\,\sigma_g \quad (4) \\
l_{base} &= \mu_{base} - 2\,\sigma_{base}, & u_{base} &= \mu_{base} + 2\,\sigma_{base}
\end{aligned}
$$

Ellipsoidal cluster: Skin color is defined to be located in an ellipsoidal cluster in NRGB. The calculation of the ellipsoid bases on the Mahalanobis distance m_p between μ and the color c_p to be classified, see (5). Since we take μ and S from the image specific skin color model, that approach adapts the classifier to the image conditions. If m_p is smaller than a threshold value t, the c_p is treated to be skin color. Empirical results show that reasonable values for t vary between 6 and 25. The evaluation is performed with $t = 9.8$.

$$
m_p = (c_p - \mu)^T \, S \, (c_p - \mu) \qquad (5)
$$

Figure 3: Classification results: (left) original image, (middle) fixed ellipsoid, (right) adaptive ellipsoid

5. Experimental results

We evaluate the parametric classifiers using the Boston University skin color database [4] and compare three ways of adaptation: (a) no adaptation, (b) automatic adaptation via the image specific skin color model, and (c) optimal adaptation. The parameters of (a) are fixed and chosen such that they are optimal for the entire set of images. The parameters of (c) are chosen such that they are optimal for each single image. Those optimal values are obtained via the

ground truth. Table 2 shows the accuracy for correctly classifying the skin color pixels (skin) and the non-skin color pixels (bg) in percent. detC denotes the determinant of the confusion matrix which measures the accuracy of the classifier. One can clearly see the increase of accuracy between (a) and (b). The result of (c) stands for the upper limit that can be achieved by that kind of classifier. Figure 3 illustrates some classification results of our experiments. Note the correct classification of facial parts such as lips and brows as non-skin colored objects. See also the excellent results in the cases of poor illumination and colored persons.

6. Conclusion and outlook

Depending on the person and the context conditions the color of human skin appears differently in each image. This makes automatic skin color classification a hard challenge. Within one image those conditions are stable and the skin color pixels look similarly. We extend a face detector with a skin color mask in order to autonomously obtain an image specific skin color model. Parametric skin color classifiers use that skin color model to adapt to the processed image, which increases the accuracy. We evaluate the increase of accuracy using a well-known image database for skin color detection. It is robust towards bad illumination and colored people. Furthermore, facial parts such as eyes, brows, lips, and teeth are detected correctly as non-skin colored objects.

We are currently extending our approach towards classifying lip color and eyebrow color and we will integrate more sophisticated classifiers.

7. References

[1] A. Broadhurst, S. Baker, "Setting Low-Level Vision Parameters", *CMU-RI-TR-04-20*, 2004

[2] S. Fischer, S. Döring, M. Wimmer, A. Krummheuer. "Experiences with an Emotional Sales Agent", *Workshop on Affective Dialogue Systems (ADS04)*, S. 309-312, Kloster Irsee, June 2004. In E. Andre et al. (eds.): ADS, LNCS 3068. Springer-Verlag

[3] S. Hämmerle, M. Wimmer, B. Radig, M. Beetz. "Sensor-based Situated, Individualized, and Personalized Interaction in Smart Environments", *Workshop für Situierung, Individualisierung und Personalisierung, 35. Jahrestagung der Gesellschaft für Informatik*, S. 261-265, Bonn, Sep. 2005. In A. B. Cremers et al. (eds.): Informatik 2005, LNI Proceedings, Volume P-67, Band 1.

[4] L. Sigal, S. Sclaroff, V. Athitsos. "Skin Color-Based Video Segmentation under Time-Varying Illumination", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, 862-877. 2004.

[5] M. Soriano, S. Huovinen, B. Martinkauppi, M. Laaksonen. "Skin Detection in Video under Changing Illumination Conditions", *In Proc. 15th International Conference on Pattern Recognition*, vol. 1, 839–842. 2000.

[6] V. Vezhnevets, V. Sazonov, A. Andreeva. "A Survey on Pixel-Based Skin Color Detection Techniques", Graphics and Media Laboratory, Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia. 2003.

[7] P. Viola, M. Jones. "Rapid Object Detection Using a Boosted Cascade of Simple Features", *CVPR*, 2001.

[8] J. Z. Wang, J. Li, G. Wiederhold, O. Firschein. "System for Screening Objectionable Images", *Computer Communications Journal 21*, pp 1355-1360, Elsevier Science, 1998.

Table 2: Comparing the results of different skin color classifiers.

	(a)						(b)						(c)					
	fixed cuboid			fixed ellipsoid			adaptive cuboid			adaptive ellipsoid			optimal cuboid			optimal ellipsoid		
seq	skin	bg	detC	skin	bg	detC	skin	bg	detC	skin	bg	detC	skin	bg	detC	skin	bg	detC
2	81.9	86.3	68.3	99.6	27.3	26.9	88.9	90.9	79.8	85.2	97.6	82.8	83.3	94.4	85.6	92.1	94.3	86.4
4	45.5	89.2	34.7	95.3	38.2	33.5	88.4	89.8	78.2	81.4	96.4	77.8	59.4	95.9	85.3	94.6	89.4	84.0
6	80.5	78.6	59.0	95.7	34.1	29.7	94.8	94.6	89.5	94.6	93.8	88.4	96.0	98.6	94.7	95.2	98.1	93.4
7	72.5	93.7	66.2	94.9	48.3	43.2	89.6	88.1	77.7	85.0	88.1	73.0	96.0	88.4	84.4	87.8	86.2	74.0
8	89.8	59.7	49.6	93.6	50.3	43.9	94.4	67.6	62.0	97.0	94.2	91.2	94.7	97.0	91.8	97.3	96.9	94.2
9	77.5	99.0	76.5	94.6	51.7	46.2	95.1	93.9	89.0	96.2	89.7	85.9	98.6	99.1	97.7	98.0	98.4	96.4
10	60.2	28.4	-11.4	93.6	57.4	51.0	98.2	44.2	42.4	98.6	54.9	53.5	68.8	96.0	64.7	89.8	90.8	80.6
11	6.0	99.2	5.2	87.2	60.8	48.0	53.6	99.9	53.5	79.0	99.7	78.7	94.4	99.2	93.6	92.4	98.1	90.5
15	96.4	43.4	39.9	84.5	63.1	47.6	97.6	45.4	42.9	96.8	76.5	73.2	94.2	96.7	91.0	91.2	96.0	87.2
16	92.3	95.3	87.6	84.8	61.6	46.3	94.6	83.7	78.3	93.6	78.5	72.2	97.4	97.6	95.0	86.0	96.6	82.6
18	97.1	99.6	96.7	84.7	60.9	45.6	99.4	95.1	94.5	97.5	94.5	92.0	98.7	99.2	97.9	97.5	94.8	92.3
21	83.5	68.2	51.7	85.5	60.4	46.0	92.7	70.4	63.1	96.4	70.9	67.3	96.3	98.0	94.3	87.5	90.7	78.2
avg	73.6	78.4	52.0	91.2	51.2	42.3	90.6	80.3	70.9	91.8	86.2	78.0	89.8	96.7	89.7	92.5	94.2	86.6