SKIN DETECTION USING SPATIAL ANALYSIS WITH ADAPTIVE SEED

Michal Kawulok∗, Jolanta Kawulok†, Jakub Nalepa, Maciej Papiez

Institute of Informatics
Silesian University of Technology
Akademicka 16, 44-100 Gliwice, Poland

ABSTRACT

This paper introduces a new method for adaptive skin detection in color images combined with spatial analysis of skin pixels. It has been reported in many works that adaptation of a skin color model to a particular image may decrease the false positives, however the false negatives are considerably high unless a local model is combined with the global one. Another possibility for improvement is to analyze spatial properties of the pixels classified as skin, but this operation strongly depends on the seed extraction technique. Our contribution lies in using a local dynamic skin model learned from the detected faces to extract seeds for the spatial analysis. We present an extensive experimental study confirming that our method outperforms alternative skin detection techniques.

Index Terms— skin detection, skin segmentation, adaptive skin color modeling

1. INTRODUCTION

Skin detection is a process, whose aim is to determine whether a given image, region or pixel presents the human skin. Its applications are of a wide range and significance, including: hand and face detection and tracking for gesture recognition and human-computer interaction, objectionable content filtering, feature extraction for content-based image retrieval, image coding using regions of interest, and many more.

There have been a number of methods proposed which adopt a pixel-wise classification scheme based on skin color modeling. Such an approach is quite effective, as it is the color which is the most discriminative feature of human skin. However, the performance of pixel-wise methods is limited due to an existing overlap in the color domain between the skin and non-skin pixels, which makes their separation impossible based on their position in a color space. The overlap may be decreased, if the model is adapted to an individual or a particular scene. There are multiple adaptation schemes which either rely on a whole-image analysis or dynamically learn the skin color model from a given skin sample. However, the local models often do not offer sufficient generalization, which increases the false negatives. Skin detection errors can also be reduced by taking advantage of the fact that skin pixels are usually clustered in the spatial domain. Hence, various region-growth operations may help reject many false positives. Here, apart from the propagation technique, the seed extraction has a critical impact on the final outcome.

Our contribution lies in combining these two aforementioned improvement directions. At first, we learn a local skin color model from detected facial regions. Afterwards, we apply this model to extract the seeds for the spatial analysis, which is performed in skin probability maps obtained using a global model. This addresses the substantial problems of adaptive skin color modeling, as well as spatial analysis of the skin pixels. The seeds extracted using a local model are highly adapted to the image, which greatly improves the outcome of the spatial analysis. Furthermore, as the latter is performed in a probability map obtained using a global model, the false negatives are not increased as in case of the local models. To the best of our knowledge, such an approach has not been explored so far, and extensive experimental study clearly confirmed the theoretical strengths of the proposed method.

The paper is organized as follows. Existing skin detection techniques are outlined in Section 2 with particular attention given to the adaptation schemes and spatial analysis methods. The details of the proposed method are presented in Section 3 and the experimental study is reported in Section 4. The paper is concluded in Section 5.

2. RELATED WORK

There are two general approaches towards skin color modeling. First, the decision rules can be explicitly defined in various color spaces to model the observed skin color distribution. There have been many such rule-based models proposed [1–4], but their effectiveness is quite low compared with the skin detectors based on machine learning. Here, the Bayesian classifier is commonly used [5,6], which is learned from the skin $C_s$ and non-skin $C_{ns}$ color histograms using two or three dimensions of various color spaces.

From the distributions, the conditional probabilities $P(v|C_s)$ and $P(v|C_{ns})$ are obtained, where $v$ indicates the color value (i.e. a $2D$ or $3D$ vector in the color space), and the final skin probability $P(C_s|v)$ is computed using the Bayes rule. In this way, any given image can be converted into a skin probability map. This method is quite effective, provided that the training set is sufficiently large.

Skin color can be modeled even from small training sets using Gaussian mixtures [7,8]. Other popular learning schemes used for skin detection include artificial neural networks [9], and support vector machines [10,11]. An interesting, thorough survey on skin color modeling was presented in 2007 by Kakumanu et al. [12].

Skin color models can be adapted to a presented image based on whole-image analysis [13–17] or from a skin sample acquired using face [18–20] or hand [21] detectors. Alternatively, skin blobs detected using a global model may be tracked in video sequences, and the skin color model can be adapted dynamically to accommodate to changing lighting conditions [22,23]. A local model is often learned using a single Gaussian [16,18] or a Gaussian mixture [14], but the
rule-based methods were also used for the adaptation [15, 21]. Also, some methods adapt the acceptance thresholds [17, 20] applied to the probability maps. In our earlier research, we used the Bayesian classifier to learn a local model from facial regions, reducing the number of histogram bins to 16 per channel in order to provide better generalization [19]. To build a Bayesian classifier from the local distribution, it is assumed that the number of non-skin pixels is complementary to the number of pixels in the skin class for each color.

Skin detection using a local model often produces high false negatives, because of low representativeness of small skin samples. This problem can be addressed by combining the local model with a global one [13, 14, 19]. Here, the final probability $P_f(C_s|v)$ can be computed as a weighted mean of the probabilities obtained using the local $P_l(C_s|v)$ and global $P_g(C_s|v)$ models:

$$P_f(C_s|v) = \omega_l \cdot P_l(C_s|v) + (1 - \omega_l) \cdot P_g(C_s|v),$$

where $\omega_l$ is the weight of the local model. Another approach adopted here consists in using a global skin color locus which imposes a restriction for the adaptation [18, 22].

Recently, Yogarajah et al. [20] proposed to adapt the thresholds in an error signal introduced by Cheddad et al. [4]. In the latter, the RGB color space is converted into the error signal $e(x)$, in which skin pixels are claimed to form a single Gaussian. Taking advantage of this observation, global thresholds in $e(x)$ were defined based on the Gaussian’s mean and standard deviation. Yogarajah’s method analyzes the distribution of $e(x)$ in a facial region to determine the decision thresholds from the obtained Gaussian’s parameters.

Spatial alignment of skin-tone pixels can be taken into account to increase the precision of determined skin region boundaries. This can be achieved using a threshold hysteresis [13] in a skin probability map. First, the seeds are extracted – they are formed by those pixels, whose skin probability exceeds a certain high threshold ($P_b$). Then, the skin regions are built from the seeds by adjoining the neighboring pixels, whose skin probability is over a threshold $P_s$, where $P_b > P_s$. Moreover, this region-growth operation is used for the skin model adaptation. The Bayesian skin model is updated based on the adjoined pixels and the new model is applied to the next frame in a sequence. This is one of few approaches to combine spatial and adaptive skin detection. Another was proposed recently [24], where the graph-cut segmentation is initialized with the facial region.

The controlled diffusion method [25] extends the threshold hysteresis with an additional condition. Namely, the distance between a source pixel $x$ and a pixel $y$ (which is to be adjoined) in the diffusion domain (i.e. color or skin probability) must be below a given threshold. During our earlier works [26, 27] we presented that using a distance transform in a combined domain of hue, luminance and skin probability outperforms the controlled diffusion technique, and therefore it has been used in the work reported here. Spatial analysis can also be performed using cellular automata [28], hidden Markov models [29] or edge detectors [30].

3. ADAPTIVE SEED FOR SPATIAL ANALYSIS

In the research reported here, our aim was to develop a scheme that would combine the benefits of the spatial analysis with the advantages of adaptive skin modeling. We observed that the characteristics of probability maps obtained using local models fit well into the expectations towards propagation seeds for the spatial analysis. Namely, local models usually make it possible to reduce the false positives, but many skin pixels remain undetected. However, as the individual's skin color distribution is similar for different visible skin parts, the false negative rate is also similar in each particular skin region. This is a very beneficial feature for the spatial analysis – if every skin region is represented with a seed, then the propagation is likely to determine correct skin boundaries.

A general flowchart of the proposed method is presented in Fig. 1. First, the faces are detected in the input image, and a local skin color model is learned from the determined facial regions. The local model is used for skin detection in the input image, and the seeds are obtained using a high-probability threshold applied in the local skin probability map. Finally, the spatial analysis from the seeds is performed in the global probability map obtained using the global skin color model. In this way, the adaptation consists in the adaptive seed detection, which dramatically improves an outcome of the spatial analysis.

Different methods can be used at each step of the proposed framework. Here, we detect faces using our face detector [31], which offers very high precision of eye localization. Skin regions are rendered as trapeze shapes [19], excluding the ellipses around the eyes. Examples of the facial regions are presented in Fig. 2c. We learned the local skin color model using two techniques, namely: i) RGB color histogram analysis (termed local histogram model) and ii) a single Gaussian (termed local Gaussian model). Using the former, the skin probability is obtained directly from the skin distribution, i.e. $P(v|C_s)$, without any assumptions on the background color distribution. Here, we decreased the number of histogram bins to 16 per channel, as justified in [19]. Following the second tech-
nique, we used a single multivariate Gaussian, which is sufficient to model local skin distributions [16, 18]. Skin probability for a given color value \( v \) is obtained as:

\[
P(v) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left[ -0.5(v - \mu)^T \Sigma^{-1} (v - \mu) \right],
\]

where \( \Sigma \) is a \( 3 \times 3 \) covariance matrix and \( \mu \) is the mean color value in the RGB color space.

Afterwards, the seeds are obtained from a binarized local probability map. In order to decrease false positives in the seeds, we apply a size-based verification procedure – those seeds, whose area is smaller than 10\% of the largest seed, are rejected. Finally, various spatial analysis schemes can be exploited, such as the threshold hysteresis [13], controlled diffusion [25], or the distance transform in a combined domain of hue, luminance and skin probability [27], introduced during our earlier works. An example of a processed image, showing the partial results at different processing steps, is given in Fig. 2. It can be seen that the final probability map (f) offers much higher discrimination between the skin and non-skin pixels (g) than the global (b, h) or local (d) probability maps.

4. EXPERIMENTAL VALIDATION

The experiments were carried out using 4000 images from the benchmark ECU database [17], acquired in uncontrolled lighting conditions. The images are associated with ground-truth binary masks that indicate skin regions. The data set was split into two equinumerous sets \(^1\). The first set was used for training the Bayesian classifier [5] (i.e. the global skin color model), while the second was used for validation. Skin detection performance is assessed based on: i) false positive rate (\( \delta_{fp} \)), i.e. a percentage of background pixels misclassified as skin, and ii) false negative rate (\( \delta_{fn} \)), i.e. a percentage of undetected skin pixels. Mutual relation of these two errors is presented using receiver operating characteristics (ROC).

The experiments were conducted using a computer equipped with an Intel Core2 Duo 2.0 GHz (4 GB RAM) processor.

Fig. 3 presents ROC curves obtained without any spatial analysis using the following schemes: i) global modeling (i.e. no adaptation), ii) local modeling (the model is trained independently for every image), and iii) adapted model (the proposed scheme). The values of \( \delta_{fp} \) and \( \delta_{fn} \) are presented in Tab. 1 (here, the acceptance thresholds were set to minimize the sum \( \delta_{fp} + \delta_{fn} \)). It can be observed that the local Gaussian model reduces the errors compared with the global model. Contrary to that, the local histogram model deteriorates the skin detection score. However, when the local models are combined with the global model using \( \omega_l = 0.5 \) in (1), then their performance is similar, while much better than the local Gaussian (see the Gaussian-based adaptation and adapted Bayesian classifier curves). Furthermore, we present the results obtained using the Yogarajah’s adaptation [20] of the thresholds applied to the Cheddad’s method [4]. Both detectors deliver a binary decision, and the obtained results are marked with asterisks in the graph in Fig. 3. For better evaluation, we extended these methods, so that their response is continuous – skin probability is inversely proportional to a distance from the middle value of the skin interval in the \( e(x) \) signal, and it is normalized, so as the probability of 0.5 is achieved at the threshold values. It can be seen that the adaptation decreases the false negatives, however the false positives slightly grow. From the ROC curves it can be seen that the adaptation offers some improvement for false positives higher than ca. 15\%, but both methods are outperformed even by the global Bayesian model.

We investigated two different techniques for spatial analysis, namely: i) threshold hysteresis [13] and ii) distance transform in a combined domain of hue, luminance and skin probability [27]. ROC curves for the threshold hysteresis are given in Fig. 4 – subsequent points on the curves were obtained using different values of \( P_{\beta} \). For the high-probability seed, \( P_{\alpha} = 0.7 \) was used as proposed in [27]. When the local model is used for seeds detection, then the errors are significantly smaller, and a slightly better result is obtained using the Gaussian-based adaptation (here we used \( P_{\alpha} = 0.9 \) for seed detection from local probability maps). The ROC curves for the distance transform are rendered in Fig. 5. A general tendency is similar to the threshold hysteresis, but the error reduction is definitely larger using the proposed scheme. We have also tried to use the Yogarajah’s adaptation:\n
\[ \delta_{fp} \quad \delta_{fn} \]

\begin{tabular}{|c|c|c|}
\hline
Method & \( \delta_{fp} \) & \( \delta_{fn} \) \\
\hline
Global Bayesian classifier [5] & 13.52\% & 10.73\% \\
Yogarajah’s adaptation [20] & 26.31\% & 7.70\% \\
Local Gaussian model & 11.49\% & 10.96\% \\
Local histogram model & 10.54\% & 15.56\% \\
Gaussian-based adaptation & 10.39\% & 8.02\% \\
Adapted Bayesian classifier [19] & 10.88\% & 8.15\% \\
Thr. hyst. (high-prob. seed) [13] & 12.12\% & 10.21\% \\
Thr. hyst. (Yogarajah’s seed) & 11.12\% & 13.48\% \\
Thr. hyst. (seed from local Gaussian) & 9.27\% & 8.13\% \\
Thr. hyst. (seed from local histogram) & 9.72\% & 8.43\% \\
Dist. transf. (high-prob. seed) [27] & 16.74\% & 5.90\% \\
Dist. transf. (Yogarajah’s seed) & 12.23\% & 12.22\% \\
Dist. transf. (seed from local Gaussian) & 8.58\% & 5.30\% \\
Dist. transf. (seed from local histogram) & 8.26\% & 5.96\% \\
\hline
\end{tabular}

---

1 See http://sun.aei.polsl.pl/~mkawulok/1cip2013
jah’s adaptation technique [20] for seed detection, but the obtained results were rather poor which can be seen from the presented ROC curves. The processing time is slightly shorter for the histogram-based seeds (ca. 1.5 seconds) than for the Gaussian-based seeds (ca. 2.2 seconds). These times may appear long, but it is sufficient to adapt the model once for a video sequence, and then the processing consists in performing the distance transform. Its processing time is ca. 0.2 seconds [27], thus allowing for almost real-time processing.

Some qualitative results are given in Fig. 6 (see http://sun.aei.polsl.pl/~mkawulok/icicp2013 for more examples). It can be seen that the proposed method offers precise segmentation outcome with very small errors (here the local Gaussian model is used as a seed for the distance transform). We have not observed the cases, where the proposed method would deliver significantly worse result than the investigated alternative techniques.

5. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a new method for skin detection in color images which combines an adaptive approach with the spatial analysis of skin pixels. Presented experimental results proved our approach to be very competitive and outperforming alternative state-of-the-art algorithms. The results also indicate how important the seed selection is for the outcome of the spatial analysis. It is worth noting that the adaptation is used only to locate the seeds, while the propagation is carried out relying on the global model. This, however, offers a significant improvement reported in the paper.

The main limitation of the presented method is that it heavily depends on the skin sample extraction that is used for the adaptation. Here we used a frontal face detector, however skin samples can also be acquired relying on a different method. Another shortcoming of the presented study is that a single local model is learned regardless of how many individuals appear in the image, and we will investigate this problem during our future work. Our ongoing research is aimed at incorporating whole-image analysis adaptation schemes in the proposed framework. Furthermore, we intent to investigate an iterative adaptation scheme, in which not only would the spatial analysis take advantage of the adaptation, but also the latter would benefit from the spatial features.
6. REFERENCES


