Human Skin Detection via Semantic Constraint

Article · July 2014

DOI: 10.1145/2632856.2632885

6 authors, including:

Changqing Zhang
Tianjin University

Jingjing Chen
City University of Hong Kong

Xiaochun Cao
Chinese Academy of Sciences

Human Skin Detection via Semantic Constraint

Binbin Ma
Tianjin University
mabinbin0202@gmail.com

Changqing Zhang
Tianjin University
zhangchangqing@tju.edu.cn

Jingjing Chen
City University of Hong Kong
chenjingjing@gmail.com

Ri Qu
Tianjin University
quri@tju.edu.cn

Jiangjian Xiao
Chinese Academy of Sciences
xiaojj@nimte.ac.cn

Xiaochun Cao✉
Chinese Academy of Sciences
caoxiaochun@iie.ac.cn

ABSTRACT

Human skin detection is a fundamental preprocessing step in many content based image processing applications and has received much attention. In this paper, we propose a novel Semantically Constrained Skin Detection (SCSD) method. Different from traditional skin detection algorithms, our method introduces the semantic constraint, which is based on the dependence between skin pixels and human body parts (skin pixels should be overlapped with body parts) to refine the detection. By employing the semantic constraint, the environmental skin-like pixels are removed effectively, while the true skin pixels constituting body parts are retained still. Experimental results on two public datasets demonstrate the significantly improved performance of our method over the state-of-the-art algorithms.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis; I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

Theory

Keywords

skin detection, semantic constraint

1. INTRODUCTION

Human skin detection aims to detect skin pixels from images, which is regarded as a binary classification problem. It has a wide range of applications in both human computer interaction and content based analysis, such as detecting and tracking of human body parts [1], face detection [15], naked people detection [9], people retrieval in multimedia databases [3], blocking objectionable content [12]. However, skin detection is laborious and very challenging. The major difficulties in skin detection are caused by various effects such as varying illumination, camera characteristics, ethnicity, individual characteristics and other factors like makeup, hairstyle, glasses, sweat and background colors [6]. Additionally, the environmental noises should be circumvented as well in practical and reliable applications.

ICIMCS'14, July. 10–12, 2013, Xiamen, Fujian, China.

(a) Input image (b) Result from [7] (c) Our result

Figure 1: Visual improvement of our approach compared with state-of-the-art method, and we marked the false positive with white circles. (a) Input image. (b) Result from [7]. (c) Our result.

In recent years, the research on skin detection mainly focuses on developing various skin models. There are many skin models that have been proposed for describing skin pixels. Generally, existing skin detection approaches can be grouped into three basic categories: physical based [11], nonparametric [13] and parametric [7] approaches. **The physical based approaches** usually rely on thresholds of different color space coordinates. **The nonparametric approaches** use a Gaussian color distribution since they assume that skins can be modeled by a Gaussian probability density function. **The parametric approaches** estimate the skin color from the trained classifier which is obtained by the training data.

Although the aforementioned methods have achieved significant success in many applications, they all suffer from the influence of background skin-like pixels. Fig. 1(b) shows an example of false positive detection results of the current approach. As shown in Fig. 1(b), the performance of skin pixels detection in body parts is accurate. But without considering about the dependence of skin pixels and body parts, the skin-like pixels are classified into skin. For example, the result from [7] detects some patches of mountain as skin, which the color of these patches is similar to the skin. In contrast, benefiting from the semantic constraint, our result in Fig. 1(c) suppresses the influence of non-body parts' skin-like pixels effectively.

Basing on the above observation, we proposed the Semantically Constrained Skin Detection (SCSD) method that uses the semantic constraint to limit the influence of background skin-like pixels. This semantic constraint is based on the

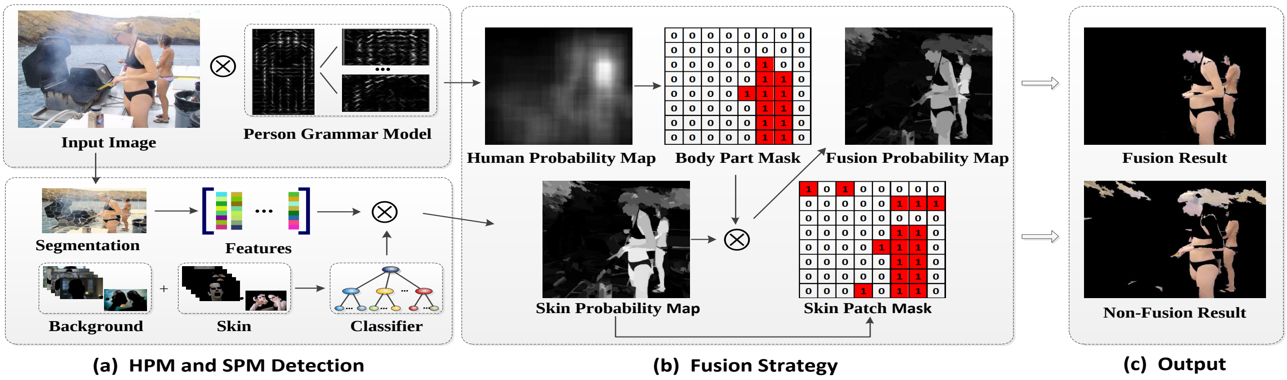

Figure 2: The flow chart of our proposed approach. (a) HPM (human probability map) and SPM (skin probability map) detection part. (b) Fusion strategy part. (c) Output image part.

motivation that skin pixels should be overlapped with body parts. In our method, The human probability map (HPM) is obtained by Cascade-DPM denoting the location of body parts. And then patch based skin detection with random forest classifier is applied to obtain the skin probability map (SPM). Finally, a fusion strategy of the two maps is proposed to perform the SCSD method. From the Fig. 1(c), we can notice that most of the background skin-like pixels are removed by our method.

The remainder of this paper is organized as follows. Section 2 describes the SCSD method in detail. The experimental results and analysises are shown in Section 3. Finally, the conclusion is drawn in Section 4.

2. THE PROPOSED APPROACH

The motivation of our method is that if we know where the person is, then the skin pixels can be detected more accurately. Therefore, in our method, we incorporate the probability map of humans into the process of skin detection. To the best of our knowledge, this is the first attempt that introduces the location information of human to detect skin in color images. As the framework shown in Fig. 2, our method mainly consists of three parts: Human Probability Map (HPM) detection, Skin Probability Map (SPM) detection and the fusion procedure of these two maps. The details will be presented in the following subsections.

2.1 HPM Detection

The probability map of human location is vital for skin detection as it can help to reduce false positive of the non-body parts. Therefore, in our approach, Cascade-DPM with Person Grammar Model [5] is employed to obtain the HPM.

Cascade-DPM is a cascade classifier from part-based deformable models, which introduce the notion of probably approximately admissible (PAA) thresholds to provide theoretical guarantees on the performance of the cascade method. Such thresholds provide theoretical guarantees on the performance of the cascade method and can be computed from a small sample of positive examples. With the consideration of these advantages, we employed the technique into our method. Let M be a model with root part v_0 and n additional parts $v_1, \cdots, v_n$, and Ω be a space of locations for each part within an image, $\triangle$ be a space of displacement. Then the $\oplus$: $\Omega \times \triangle$ be a binary operation taking a location and a displacement to a "displaced location". For the human model, the training process is solving the following formulation:

$$\text{score}(\omega) = m_0(\omega) + \sum_{i=1}^{n} \max_{\delta_i \in \triangle}(m_i(a_i(\omega) \oplus \delta_i) - d_i(\delta_i)), \quad (1)$$

where the $m_i(\omega)$ is the score for placing v_i in location ω, the $a_i(\omega)$ specify the ideal location for v_i as a function of the root location and $d_i(\delta_i)$ specify a deformation cost for a displacement of v_i from its ideal location relation relative to the root. Finally, a cascade detection algorithm is outlined for a general class of models defined by a grammar formalism. This class includes not only tree-structured pictorial structures but also richer models that can represent each part recursively as a mixture of other parts.

2.2 SPM Detection

Now, we focus on the SPM. To get a better performance, edge-based color constant [14] method is employed to ensure the color constancy of the original image. In addition, two types of descriptors, covariance descriptor and tamura texture descriptor, are combined to enhance the detection results.

We get the SPM with Random Forests classifier [2], as shown in Fig. 2. In our framework, we choice the classifier for the reason that random classifier can achieve the best performance [2]. Random Forests is an ensemble of tree predictors such that each tree depends on the values of a random vector. This vector is sampled independently from the same distribution for all the numerous trees in the forest.To classify a new object from an input vector, the input vector is presented to each of the trees in the forest. And each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes.

In our method, let $\mathbf{X}$=$(\mathbf{x}_1, ..., \mathbf{x}_n)^T \in \mathcal{R}^{n \times d}$ denote the training data set, which consists of n patches in d-dimensional feature space. And the $\mathbf{X}$ are labeled by $\mathbf{y} = (y_1,..., y_n)$ where $y_i \in \{0, 1\}$ represents the binary class label assigned to $\mathbf{x}_i$. Then the $\mathbf{X}$ and $\mathbf{y}$ are sampled for growing classify trees. For the kth tree, a random vector Θ_k is generated for sampling. Then we get an ensemble of classifiers $h_1(\mathbf{x}, \Theta_1)$,

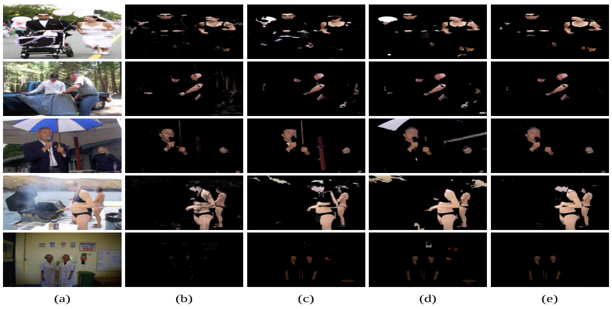

Figure 3: Visual comparison of different approaches in *ETHZ PASCAL dataset*. (a) Original images. (b) The result of Mul-Model [8]. (c) The result of J48-Sys [7]. (d) The results of SD. (e) The result of SCSD.

Table 1: Comparison Between Different Approaches on *Sigal dataset*

Approaches	Accuracy	Precision	Recall	F-measure
Mul-Model	0.846	0.667	0.695	0.681
J48-Sys	0.877	0.738	0.740	0.739
SD	0.864	0.708	0.720	0.714
SCSD	**0.902**	**0.790**	**0.797**	**0.793**

$h_2(\mathbf{x},\Theta_2)$,..., $h_k(\mathbf{x},\Theta_m)$ as follow:

$$\mathrm{H}(\mathbf{x}) = \arg\max_y \sum_{i=1}^{m} \mathbf{I}(h_i(\mathbf{x},\Theta_i) = y), \quad (2)$$

where $\mathrm{H}(\mathbf{x})$ is the combined classification model, $I(\cdot)$ is the indicator function and y is a label variable.

2.3 Fusion Strategy

In this section, we will state our fusion strategy which is effective in increasing the accuracy of skin detection. This fusion method combines the human probability map and skin probability map effectively by using the product rule.

To serve as our fusion part, we first apply the common combination strategies to skin detection. Then we discuss their performance to motivate our fusion strategy. Let I, C_{skin} and G_{human} denote the input image, skin color classifier and person grammar model respectively. And let $D_{spm}(I, C_{skin})$ and $D_{hpm}(I, G_{human})$ denote the probability matrices from SPM and HPM, respectively. The fusion rules can be obtained as follow:

$$\begin{aligned}\mathrm{D}(p) &= P(y_p = 1|D_{spm}(I, C_{skin}), D_{hpm}(I, G_{human})) \\ &\propto \frac{1}{Z}(\zeta(D_{spm}(I, C_{skin})) + \zeta(D_{hpm}(I, G_{human})), \end{aligned} \quad (3)$$

where $\mathrm{D}(p)$ represents the combined results value of pixel p, y_p is a binary random variable taking the value 1 if p is a salient pixel and 0 otherwise, and Z is a constant for normalization. We implemented three different options for the function ζ in Equation 3, including

$$\zeta_1(x) = x,\ \zeta_2(x) = \exp(x),\ \text{and}\ \zeta_3(x) = \frac{-1}{\log(x)}. \quad (4)$$

We used these standard aggregation methods to combine the HPM and SPM. However, the results is far from the expectation. The main reasons are that they do not consider the different functions among different methods and treat them equally. Actually the values of HPM are not the probability of skin as in SPM. Therefore, we propose our fusion strategy as follow.

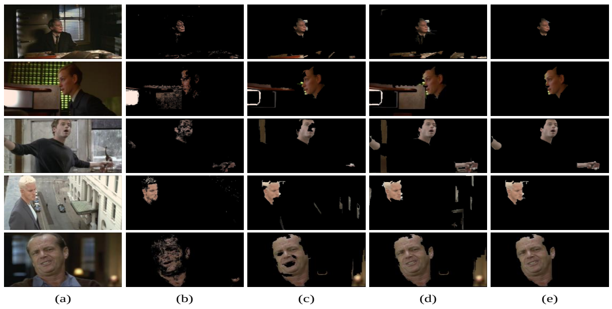

Figure 4: Visual comparison of different approaches in *Sigal dataset*. (a) Original images. (b) The result of Mul-Model [8]. (c) The result of J48-Sys [7]. (d) The results of SD. (e) The result of SCSD.

Table 2: Comparison Between Different Approaches on *ETHZ PASCAL dataset*

Approaches	Accuracy	Precision	Recall	F-measure
Mul-Model	0.817	0.573	0.633	0.602
J48-Sys	0.834	0.607	0.679	0.641
SD	0.826	0.590	0.661	0.624
SCSD	**0.852**	**0.646**	**0.711**	**0.677**

$$\mathrm{D}(p) = \begin{cases} s_p \times \lambda_1, & \text{if } h_p > l \\ s_p \times \lambda_2, & \text{if } h_p \le l \end{cases} \quad (5)$$

where h_p and s_p are values in the Human Probability Matrix and Skin Probability Matrix, respectively, l is the detection threshold of human body parts and λ_1, λ_2 are reweighing parameters that enlarge the probability of skin in body parts and reduce the the probability of skin in non-body parts. And we will describe the settings of l, λ_1 and λ_2 in the following part.

3. EXPERIMENT

For experimental evaluation, we use two datasets to show the the comparison between state-of-the-art approaches and ours both qualitatively and quantitatively. And we select 3 methods as the comparisons on 4 evaluation metrics.

3.1 Datasets and Experiment Settings

Datasets: We conduct experiments on two dataset: Sigal dataset [10] and ETHZ PASCAL dataset [4]. Brief introductions about these two datasets are as follows. Sigal dataset contains a set of 21 high quality sequences from nine popular movies. Images in ETHZ PASCAL dataset are sampled from PASCAL VOC 2009 and the total size of the dataset is 549. Skin detection on these two dataset is challenging, because images from these two datasets span a wide range of environmental conditions, such as different ethnicity, skin tones and lighting conditions. Some scenes contain multiple people and multiple visible body parts. These images contain both indoor scenes and outdoor scenes, with static and moving camera. The lighting varies from natural light to directional stage lighting. Some images contain shadows and minor occlusions.

Experiment Settings: In our method the body parts detection threshold l and reweigh parameters λ_1, λ_2 are set as 0.56, 1.2 and 0.85 respectively. For performance comparison of our approach, we select a physical based approach (Mul-Model) [8], a parametric approach (J48-Sys [7]) and our method without semantic constraint (SD) as the comparisons. We evaluate the performance based on Accuracy, Precision, Recall and F-measure.

3.2 Results and Analysis

Experimental results on Sigal dataset: We first conduct our experiments on Sigal dataset. Fig. 4 shows the experiment results. From Fig. 4, it is obvious that the non-body parts skin-like pixels are reduced and the true skin pixels are detected more accurately. The reason is that we take into account the dependence between body part and skin pixels, which can help remove the low probability skin like patches and strengthens the probability of body parts. Take the second input image ($2nd$ row) in Fig. 4 for example, other methods tend to mistake the desk and screen into skin due to the similarity of color among desk, screen and skin, while our method can classify the skin correctly by reducing the weight of non-body parts pixels. Meanwhile, our method detects the skin pixels of body parts accurately by enlarging the probability of body parts pixels. For instance, other methods tend to classify the shadows on the neck into non-skin due to the influence of shadows, while our method tends to classify it into skin. Furthermore, a quantitative evaluation result is shown in Table. 1. From the Table. 1, we can find that our method achieves the highest F-measure of 0.793 and outperforms the three other methods in F-measure 7% to J48-Sys, 16% to Mul-Model and 11% to SD, respectively. And we can also find that our Accuracy, Precision and Recall all have a certain degree of improvement.

Experimental results on ETHZ PASCAL dataset: Then we conduct the experiments on ETHZ PASCAL dataset. The comparison methods are the same with Sigal dataset. Fig. 3 shows the 5 samples of experiment results. It should be notice that ETHZ PASCAL dataset is very challenging dataset, because its images contain more complex background. Besides, some images contain shadows and minor occlusions. From Fig. 3, we can see that compared with other methods, our method achieves the best performance. Take the second input image ($2nd$ row) in Fig. 3 for example, it can be seen that only our method can distinguish the skin patch from those skin-like pathes since the color of forest and wheel is similar to skin color in this image. We further compare the accuracy, precision, recall and F-measure among our method and other methods. Table. 2 shows the comparison results. From the Table. 2, we can see that our method achieves the highest score in four evaluation metrics. From the F-measure we can notice that SCSD increased detection performance of almost 5.6% to J48-Sys, 12% to Mul-Model, 8% to SD. The experimental results demonstrate that our method can obtain a satisfactory performance even in complex background.

4. CONCLUSIONS

In this paper, we propose a novel semantically constrained skin detection approach (SCSD). This semantic constraint is essentially the dependence between skin pixels and body parts. By introducing this semantic constraint, our method can suppress non-body parts false positive significantly. Besides, a fusion strategy is proposed to combine the HPM and SPM in our method. Experimental results show that our method achieves the best performance of skin detection compared to the state-of-the-art methods. In the future, we may focus on speeding up our method and realize the real-time skin color detection. In addition, we will exploit other prior knowledge to further improve the skin detection accuracy.

5. ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China (No. 61272264, 61170178, 61273276), National High-tech R&D Program of China (2014BAK11B03), and 100 Talents Programme of The Chinese Academy of Sciences.

6. REFERENCES

[1] A. A. Argyros and M. I. Lourakis. Real-time tracking of multiple skin-colored objects with a possibly moving camera. In *ECCV*, pages 368–379, 2004.

[2] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[3] L. L. Cao, X. L. Li, N. H. Yu, and Z. K. Liu. Naked people retrieval based on adaboost learning. In *ICMLC*, pages 1133–1138, 2002.

[4] M. Everingham, L. Van-Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge 2009. In *2th PASCAL Challenge Workshop*, 2009.

[5] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Cascade object detection with deformable part models. In *CVPR*, pages 2241–2248, 2010.

[6] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern recognition*, 40(3):1106–1122, 2007.

[7] R. Khan, A. Hanbury, R. Sablatnig, J. Stöttinger, F. A. Khan, and F. A. Khan. Systematic skin segmentation: merging spatial and non-spatial data. *Multimedia Tools and Applications*, pages 1–25, 2012.

[8] R. Khan, J. Stöttinger, and M. Kampel. An adaptive multiple model approach for fast content-based skin detection in on-line videos. In *ACM MM*, pages 89–96, 2008.

[9] J. S. Lee, Y. M. Kuo, P. C. Chung, E. Chen, et al. Naked image detection based on adaptive and extensible skin color model. *Pattern Recognition*, 40(8):2261–2270, 2007.

[10] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *Transactions on Pattern Analysis and Machine Intelligence*, 26(7):862–877, 2004.

[11] K. Sobottka and I. Pitas. A novel method for automatic face segmentation, facial feature extraction and tracking. *Signal processing: Image communication*, 12(3):263–281, 1998.

[12] J. Stöttinger, A. Hanbury, C. Liensberger, and R. Khan. Skin paths for contextual flagging adult videos. In *Advances in Visual Computing*, pages 303–314, 2009.

[13] H. M. Sun. Skin detection for single images using dynamic skin color modeling. *Pattern recognition*, 43(4):1413–1420, 2010.

[14] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *Transactions on Image Processing*, 16(9):2207–2214, 2007.

[15] M. H. Yang and N. Ahuja. Detecting human faces in color images. In *ICIP*, pages 127–130, 1998.