HUMAN SKIN DETECTION IN IMAGES BY MSER ANALYSIS

Lei Huang[1,2], Tian Xia[1], Yongdong Zhang[1], Shouxun Lin[1]

[1]Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2]Graduate University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

Human skin detection in images is desirable in many practical applications, e.g., adult-content filtering. However, existing methods are mainly pixel-based and ignore that human skin is region-based. In this paper, we introduce a successful region detector, i.e., MSER, into the skin detection by regarding the skin region as the maximally stable extremal region (MSER). We extend the original MSER to both color and texture analysis to reduce the skin-like regions[1]. Furthermore, to be adaptive to the dynamic illumination and chrominance, face detection is used to customize the skin color model to each image. The proposed method has achieved promising performance over our dataset, which is a challenging set with a great part of hard images. Our *True Positive Rate is* 81.2% under *False Positive Rate* 8.2%, which outperforms all of eight state-of-the-art algorithms.

Index Terms—skin detection, MSER analysis, skin distance image, texture map, Gaussian model

1. INTRODUCTION

Human skin detection in a single image plays an important role in various applications, e.g., human body detection, human-computer interaction, pornographic content filtering. However, due to the variable ambient lights, confusing backgrounds, and diversity of human races, detecting human skin region in a single image is really not an easy task.

There is plenty of previous work on human skin detection in images. Generally, the detection approaches fall into two categories: pixel-based methods and region-based ones. Pixel-based skin detection has a long history, a comprehensive survey of which is provided in [1]. Concretely, the pixel-based methods can be divided into two kinds: static methods and dynamic ones. The static methods are based on static color model, which usually utilizes fixed color thresholds [2][3][4] or a pre-trained skin distribution [5][6]. They are simple and with a good time efficiency, but are not robust to the variations of illumination and chrominance. Therefore, the dynamic methods are proposed to address the above issues. Yang et al. [7] proposed an adaptive skin-color model for tracking the human face in videos. It is based on the assumption that there is a relationship between continuous frames in the aspects of illumination condition. However, there is no correlation between single images. For a single image, Rowley et al. [8] utilized detected face region to build a Gaussian skin color model. Lee et al. [9] used five pre-trained skin chroma clusters to select the most suitable skin model. However, in the case of no skin or less skin in the five predefined area or the illumination is not consistent with the pre-trained model, this method will not work. Sun [10] divided the skin detection into training stage and detection stage. At the detection stage, preliminary skin pixels are detected to construct the dynamic skin color model.

We can see that all the pixel-based methods aim to classify each pixel as skin or non-skin individually and independently from its neighbors. It is not in line with the knowledge that human skin is patch-shaped. The region-based methods try to utilize the spatial arrangements of skin pixels to improve the accuracy. Kruppa et al. [11] proposed that each skin region should be fit to an ellipse. Although this assumption can remove some skin-like regions, but it still cannot cope with the changes of chrominance and illumination.

From the above survey, we can see that the region-based skin detection is more reasonable than pixel-based detection since skin is patch-shaped. Within the skin region, the variations of the color/texture are small. However, the color/texture variations at the boundaries of the skin region are obviously larger than the ones from the internal of skin region. Following the above analysis, we deem that the skin region can be represented as the maximally stable extremal region in MSER detector [12], which is widely used in region detection [13][14]. In [13], Donoser et al. have proposed to use MSER analysis for color blob segmentation, however, their method only focuses on color analysis and requires people to label the region-of-interest (ROI) manually for each single image. In this paper, we propose to utilize MSER analysis to detect the human skin in color

[1] Skin-like region is the region which shares similar color with human skin but is not human skin, e.g. certain types of wood.

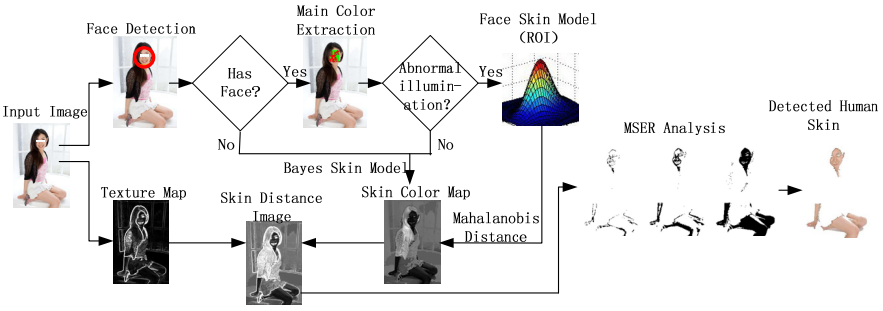

Fig. 1. Flowchart of our algorithm.

images. In order to reduce the interference of skin-like regions, we extend the original MSER to both color and texture analysis. Furthermore, to be adaptive to the dynamic illumination and chrominance, we utilize face detection to customize the skin color model to each single image.

Our contribution is that we propose a novel dynamic region-based human skin detection method through MSER analysis and we extend the original MSER to both color and texture to remove skin-like regions. Furthermore, for images with face, skin color map is constructed from face regions for invariant illumination.

2. HUMAN SKIN DETECTION BY MSER ANALYSIS

Our human skin detection method can be divided into three main steps. First, skin color model is constructed. Second, the skin distance image is generated by the combination of skin color map and texture map. Finally, a modified version of MSER algorithm is performed on the skin distance image to obtain the skin regions. The flowchart of the proposed method is shown in Fig.1.

2.1. Selection of skin color model

To be adaptive to the dynamic illumination and chrominance, we utilize face detection [15] to select the skin color model.

For images with face detected, Bayes skin color model [5] is first performed on the face region to get primary face skin pixels. Then, the main color region Ω of face is obtained by histogram analysis over the whole face region, and the normal illumination ratio $\Delta = spn / pn$ is calculated, where spn is the number of skin pixels in Ω, pn is the number of total pixels in Ω. If $\Delta < 0.3$, it implies that most of the face skin pixels do not satisfy the Bayes skin color model due to the abnormal illumination, and a Gaussian skin color model M_{gauss} [8] is built in this case. For images with more than one face, each face will derive a Gaussian model, and the final model is the linear combination of separate ones.

For images without face detected, Bayes skin color model [5] is selected here.

2.2. Generation of skin distance image

According to the characteristics of MSER, we need to construct a skin distance image in which the value of each pixel expresses the distance to real human skin. In our methods, the skin distance image is a combination of both skin color map and texture map.

Skin color map. For the face skin model, the skin color map CM is calculated by the Mahalanobis distance. The Mahalanobis distance $D_M(\vec{p})$ between a pixel $I(x,y)$ in the image I and the skin model M_{gauss} (obtained in Section 2.1) is defined in Eq. (1).

$$CM(x,y) = D_M(\vec{p}) = (\vec{p}-\vec{u})^T \Sigma^{-1} (\vec{p}-\vec{u}), \quad (1)$$

where $\vec{p}$ is a 3×1 vector with Y, C_b, C_r values of pixel $I(x,y)$, $\vec{u}$ and Σ are the mean vectors and the covariance matrix of M_{gauss} separately.

For Bayes skin color model(non-face images and face-images with normal illumination acquired in section 2.1), $CM(x,y)$ is calculated as follows:

$$CM(x,y) = \frac{1}{P(skin \mid rgb)} * 10. \quad (2)$$

$CM(x,y)$ should be normalized to [0, 255].

Texture map. With the observation that lots of the skin-like regions are coarse and the human skin regions are smooth. Therefore, we can remove the skin-like regions by texture analysis. In this step, we compute a texture map TM which can reflect the level of coarseness of the input image. We use a simple texture feature to construct the texture map TM. The texture feature $TM(x,y)$ of pixel $I(x,y)$ is calculated as follows:

$$\begin{aligned} TM(x,y) = {} & abs(I_g(x+1,y-1) - I_g(x-1,y+1)) \\ & + abs(I_g(x+1,y) - I_g(x-1,y)) \\ & + abs(I_g(x+1,y+1) - I_g(x-1,y-1)) \\ & + abs(I_g(x,y+1) - I_g(x,y-1)) \end{aligned}, \quad (3)$$

where $abs(\cdot)$ is absolute value function, I_g is the gray-level image of I.

Skin distance image. Skin distance image DI is estimated both from the skin color map CM and the texture map TM.

$$DI(x,y) = CM(x,y) + \beta * TM(x,y), \quad (4)$$

where β is the weight of the texture map, it is determined empirically in applications. In our experiment, β is set to 3.

2.3. MSER analysis

MSER analysis is carried out on the skin distance image DI, and the union of the detected MSERs is the final skin regions. According to MSER algorithm [12], the maximally stable extremal regions R is defined as:

$$\rho(R;\Delta) = \frac{|R_{+\Delta}| - |R_{-\Delta}|}{R}, \quad (5)$$

where R is a maximally stable extremal region when $\rho(R;\Delta)$ is minimum. The extremal region, also known as $\alpha - connectivity$ [16], is a maximal connected component of a level set $S(i)$ in which the pixel intensity is not greater than i. $R_{+\Delta}$ is the smallest extremal region containing R, and has intensity which exceeds of at least Δ intensity of R. $R_{-\Delta}$ is the biggest extremal region contained by R, and has intensity which is exceeded at least Δ intensity by R.

The original MSER procedure is carried out from $i = 0$ to L_m for minimum MSERs, and from $i = L_m$ to 0 for maximum MSERs, L_m is the maximum gray value of the input image. In our case, we only need the minimum MSERs. Thus, the MSER analysis is carried out from $i = 0$ to L_s on DI. For face skin model, L_s is equal to $3 * k * \sigma + \omega$; and for Bayes skin model, L_s is equal to $\phi + \omega$, where σ is the standard deviation of the skin model M_{gauss}, k, ϕ, ω are positive constants, k and ϕ are the thresholds for the skin color map component, and ω is the threshold for the texture map component.

2.4. Human skin detection for non-face images

The above sections (section2.1, 2.2) show that for face images, face has been used to customize the skin color models to the variations of illumination, however, for non-face images, Bayes skin color model [5] is selected for construction of skin color map. Thus, for non-face, our method cannot handle the variations of illumination. However, there are many applications have a large ratio of face images, e.g. human-computer interaction, pornographic images filtering. Therefore, the proposed method is promising in applications.

3. EXPERIMENTAL RESULTS

In this section, we conduct a comprehensive comparison with eight popular algorithms to evaluate the proposed approach.

3.1. Dataset

To evaluate the algorithm objectively, we build a new human skin dataset. This dataset contains 1,000 images randomly sampled from "HOT or NOT"[2] which is a popular website. We labeled the ground-truth of skin for these images by Photoshop. Our dataset contains a great part of challenging images with variable ambient lights, confusing backgrounds, diversity of human races; and also various resolutions and visual quality. It contains 38,868,720 skin pixels and 139,091,233 non-skin pixels. 71.9% images in the dataset are detected to contain face. The ratio looks a little high. However, by analyzing the images in pornographic image detection which is an important application of skin detection, we find that there are 68.1% of 17,422 pornographic images are detected to contain face. Thus, the ratio of images containing face in our dataset is consistent with real applications.

3.2. Baseline algorithms and evaluation metrics

To evaluate the effectiveness of the proposed method, we compare it to eight popular algorithms, i.e. CBCR[2], Cluster-RGB [3], YUV&YIQ [4], BAYES[5], GMM[5] , MEM[6], GAUSS [7], and Color-Shape [11]. True Positive Rate (TPR) and the False Positive Rate (FPR) are used to evaluation these algorithms.

3.3. Evaluation on dataset

Fig.2 shows the performance comparison on our dataset, in which four algorithms with probabilistic outputs are presented in ROC curves, the other five algorithms with binary outputs are presented in a point in the figure.

In real application, the performance with FPR below 10% attracts the most attention, and we can see that our method achieves the largest TPR in this range.

3.4. Varying illumination and chrominance test

To illustrate the performance of our method in the case of varying illumination and chrominance, we show some samples in Fig.3. Due to the limit of paper length, we only compare to three algorithms, i.e., BAYES, GMM, and MEM, which achieve better performance than others in the above evaluation. In Fig.3, the first row concerns abnormal chrominance, and the second row concerns abnormal illumination, we can find that our method, in red boxes, outperforms other methods under varying illumination and chrominance.

3.5. Evaluation of the effect of texture

According to Eq.(4), we can see that parameter β controls the effect of texture in MSER analysis. A larger β means the proposed method pays more attention to texture information, and $\beta = 0.0$ means only color information is considered in MSER analysis. In Fig.4, we show some samples to illustrate the skin texture map is effective to remove skin-like regions. We can also find that when β is too large (e.g. $\beta = 5.0$), there will be many skin areas missing; on the other hand, when β is too small (e.g. $\beta = 1.0$ (0.0)), the skin-like region will be falsely detected. So, in our experiments, β is set 3.0.

[2] http://www.hotornot.com

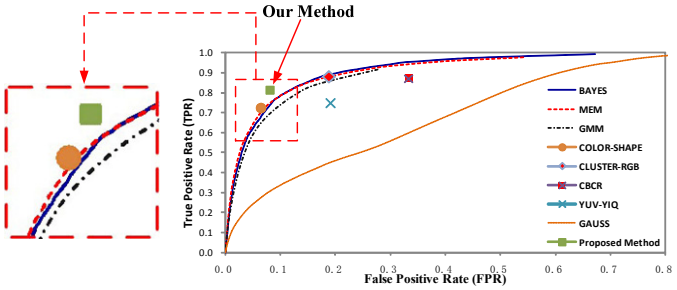

Fig. 2. Performance comparison on our dataset.

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 3. Performance comparison under different chrominance and illumination: (a) and (f) the input images; (b) and (g) the results of our method; (c) and (h) the results of BAYES; (d) and (i) the results of GMM; (e) and (j) the results of MEM.

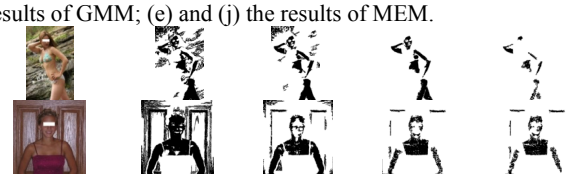

Fig. 4. Results of the proposed method with different β in Eq.(4).

4. CONCLUSIONS AND FUTURE WORK

Under the assumption that skin regions in images can be regarded as MSERs, we propose a novel dynamic region-based human skin detection method through MSER analysis. By extending the original MSER to both color and texture analysis, the skin-like regions have been effectively removed. Meanwhile the introduction of face detection for skin color model customization has greatly enhanced the robustness to dynamic illumination and chrominance for images with face. The limitation is that for non-face images, our method cannot handle variations of illumination which is a challenge task. But, with the skin color map based on Bayes skin color model and improved MSER analysis, our method also gets promising results for non-face images.

In the future, we plan to extend our algorithm from binary outputs to probabilistic outputs to fulfill different practical applications.

5. ACKNOWLEDGEMENT

This work was supported by the Beijing Natural Science Foundation under Grant 4112055; by the National Basic Research Program of China (973 Program) under Grant 2007CB311100; by the National High Technology and Research Development Program of China (863 Program) under Grant 2009AA01A403; by the Co-building Program of Beijing Municipal Education Commission.

6. REFERENCES

[1] V. Vezhnevets, et al., “A survey on pixel-based skin color detection techniques,” In 13th International Conference on the Computer Graphics and Vision, pp.85-92, 2003.

[2] R.L. Hsu, et al., "Face detection in color images," IEEE Transactions on PAMI, 24(5), pp. 696-706, 2002.

[3] P.J. Kovac, et al., “Human skin colour clustering for face detection,” The IEEE Region 8 Computer as a tool EUROCON, pp.144-148, 2003.

[4] L.J. Duan, et al., “Adult image detection method base-on skin colour model and support vector machine,” Asian Conference on Computer Vision, pp.797-800, 2002.

[5] M.J. Jones, et al., “Statistical color models with application to skin detection,” IJCV, 46(1), pp. 81-96, 2002.

[6] B. Jedynak, et al., “Maximum entropy models for skin detection”, Technical Report XIII, Universite des Sciences et Techonlogies de Lille, France, 2002.

[7] J. Yang, et al., “Skin-color modeling and adaptation,” Asian Conference on Computer Vision, pp.687-694, 1998.

[8] H.A. Rowley, et al., “Large scale image-based adult-content filtering,” VISAPP, pp.290-296, 2006.

[9] J.S. Lee, et al., “Naked image detection based on adaptive and extensible skin color model,” Pattern Recognition, 40(8), pp.2261-2270, 2007.

[10] H.M. Sun, “Skin detection for single images using dynamic skin color modeling,” Pattern Recognition, 43(4), pp.1413-1420, 2010.

[11] H. Kruppa, et al., “Skin patch detection in real-world images,” In Annual Symposium for Pattern Recognition of the DAGM, pp.109-117, 2002.

[12] J. Matas, et al., “Robust wide-baseline stereo from maximally stable extremal regions,” In Proc. BMVC, pp.384-393, 2002

[13] M. Donoser, et al., “Color blob segmentation by mser analysis,” In Proc. ICIP, pp.757-760, 2006.

[14] P.E. Forssen, “Maximally stable colour regions for recognition and matching,” In Proc. CVPR, pp. 1220-1227, 2007.

[15] P. Viola, et al. “Rapid object detection using a boosted cascade of simple features,” In Proc. CVPR, pp. 511-518, 2001.

[16] P. Soille. “Constrained connectivity for hierarchical image partitioning and simplification,” PAMI, 30 (7), pp.1132-1145, 2008.