Skin-Color Detection Based on Adaptive Thresholds

Ming-Ji Zhang[1,2], Wei-Qiang Wang[1,2], Qing-Fang Zheng[1,2], Wen Gao[1,2]
1. Institute of Computing Technology, Chinese Academy of Sciences
2. Graduate School of the Chinese Academy of Sciences
{mjzhang, wqwang, qfzheng, wgao}@jdl.ac.cn

Abstract

In this paper a new skin detection method based on adaptive thresholds is proposed. Compared with the fixed threshold histogram method used widely, ours can find optimal thresholds to the different complex backgrounds. Four clues are summarized from the skin probability distribution histogram (SPDH) to help search candidates of optimum thresholds, and an ANN classifier is trained to select the final optimum threshold. A novel image relation operation is also proposed to eliminate confusing backgrounds. The method is fast thus appropriate for real-time applications since no iterative operation is involved. Experimental results show that the proposed method can achieve better performance than the fixed threshold histogram method.

1. Introduction

Skin-color detection has been employed in many applications such as face detection, gesture recognition, human tracking, pornographic image filtering, etc. It has been proved that using the skin-color information in the pre-process can reduce the difficulties of problems. But due to the effects of different human races, ambient lights and confusing backgrounds, detecting skin-color accurately is not an easy task.

Many approaches have been proposed to detect skin-color in static images, and can be classified into three categories: explicitly defined skin region [1,2], nonparametric skin distribution modeling [3] and parametric skin distribution modeling [4]. However, these methods are all sensitive to the different illuminations and complex backgrounds, therefore some adaptive methods are also proposed recently. Huynh-Thu et al [5] proposed a method to find the optimal threshold automatically for the each sub-model in the GMM. In [6] Phung et al proposed an adaptive skin segmentation technique which employed the texture characteristics of the human skin. Cho et al [7] also presented a skin-color filter that was capable of presented a skin-color filter that was capable of adaptively adjusting its thresholds box and effectively separating skin-color regions from similar background-color regions. But all the methods involve iterative processes, thus increase the computational cost.

In this paper, we proposed a new effective skin-color detection method. In section 2, the skin probability distribution histogram (SPDH) is constructed from the original image, and four clues are summarized from training set to indicate the characters of candidates of optimum thresholds. ANN and C4.5 algorithms are compared to establish the rules to select the final optimum threshold from the candidates, while no iterative operations are needed. In section 3, an upper limit threshold and an image relation operation are introduced to eliminate confusing backgrounds. In section 4, some experimental results are discussed.

2. Adaptive Thresholding

The Skin Probability Map (SPM) [1,3,8] is a widely used method for skin color detection, which has been assessed as the best one in terms of recall and precision [8]. However, the decision threshold for SPM is fixed and can't cope with the various image conditions. In this section, a novel method of selecting adaptively the optimum threshold is proposed, and the SPDH is introduced to help the process.

2.1. Skin probability map (SPM)

After establishing the skin and non-skin histograms from training set, according to the Bayesian rules, the skin probability of a pixel with color value x is defined as follows.

$$p(\omega_{skin} \mid x) = \frac{p(x \mid \omega_{skin}) \cdot p(\omega_{skin})}{p(x \mid \omega_{skin}) \cdot p(\omega_{skin}) + p(x \mid \omega_{non-skin}) \cdot p(\omega_{non-skin})} \quad (1)$$

where ω_{skin} and $\omega_{non-skin}$ denote the classes of skin and non-skin, $p(x \mid \omega_{skin})$ and $p(x \mid \omega_{non-skin})$ are the prior probabilities of skin and non-skin which are computed

from skin and non-skin histograms. $p(\omega_{skin})$ and $p(\omega_{non-skin})$ are the class probabilities which can be estimated from the skin and non-skin histograms. A pixel x is labeled as skin pixel if

$$p(\omega_{skin} \mid x) \geq \tau \tag{2}$$

where τ is a fixed threshold for decision.

2.2. Skin probability distribution histogram

The fixed threshold, however, isn't well suited for every image since it is often a statistical compromise between true positive and false positive on the training set. An alternative is to find an optimum threshold for each image according to its content.

Our experimental observation on a large image set indicates that some clues summarized from the SPDH can help choose the optimum threshold. The SPDH is a histogram that represents the distribution of skin probabilities of all the pixels in an image. Its x-axis represents the normalized skin probability, and the normalized skin probability p_x of a certain pixel x in an image I is:

$$p_x = \frac{p(\omega_{skin} \mid x)}{\max_x \{p(\omega_{skin} \mid x) \mid x \in I\}} \tag{3}$$

The y-axis represents the standardized total count of pixels with a certain normalized skin probability:

$$S(p_x) = \frac{\sum_{y \in I} \delta(p_y - p_x)}{\sigma_\Lambda}, \quad \delta(p_y - p_x) = \begin{cases} 1, & p_y - p_x = 0 \\ 0, & otherwise \end{cases} \tag{4}$$

where σ_Λ is the standard deviation of colors of all the pixels in the set $\Lambda = \{y \mid p_y = p_x, y \in I\}$. Introducing the denominator σ_Λ to standardize the Y-coordinate is due to the following reason: through the experimental observations, we find that the SPDH of skin regions usually exhibits unimodality, which can help segment skin regions. But obviously different colors maybe have the same skin probability, which make the SPDH of some non-skin regions overlap partly with the SPDH of skin regions and destroy the unimodality of skin regions. But when σ_Λ is introduced as denominator, the Y-coordinates in overlapped intervals will shrink more (σ_Λ >1) or extend less (σ_Λ <1) than that in unoverlapped intervals. It helps to maintain the unimodality of skin regions and therefore is helpful to the selection of optimum thresholds. For the same sake, we also smooth SPDHs with a mean filter. After that, four clues can be derived from SPDHs to help choose the appropriate threshold, they are:

1. The SPDH of the skin regions of an image is usually unimodal, therefore looks like a hill shape, we call it *skin hill*, and it generally lies in the interval of high normalized skin probabilities.
2. The SPDH of non-skin regions generally concentrate in the interval of low normalized skin probabilities.
3. The summit of the skin hill usually overlaps with the summit on the right end of the SPDH.
4. The optimum threshold always lies on the left foot of the skin hill, which is also one of the valleys of a SPDH.

These clues are intuitively comprehensible. Skin pixels generally have high skin probabilities, therefore fall in the bins on the right end of the SPDH, while non-skin pixels generally fall in the bins on the left end. Although the skin-colors of a man are uniform theoretically, they actually exhibit a normal distribution in an image because of the light reflection on the surface of human bodies. Thus it's easy to understand hill-shape of skin regions in the SPDH. In Table 1, 100 images which contains skin are observed to verify the validity of the four clues.

Table 1. Statistical data for the validities of the four clues

Total number	Clue 1	Clue 2	Clue 3	Clue 4
100	87	76	91	93

2.3. Finding optimum threshold

Based on these clues, only the valley of the SPDH may be the optimum threshold, therefore all the valleys are regard as the candidates. In most conditions, the optimum threshold can be easily found (fig 1.a), but the possible effects of different illuminations and confusing backgrounds have to be considered. In this condition, the distributions of skin and non-skin regions will overlap partly (fig 1.b), which violates some clues of the above. Therefore some robust rules have to be established to cope with those situations.

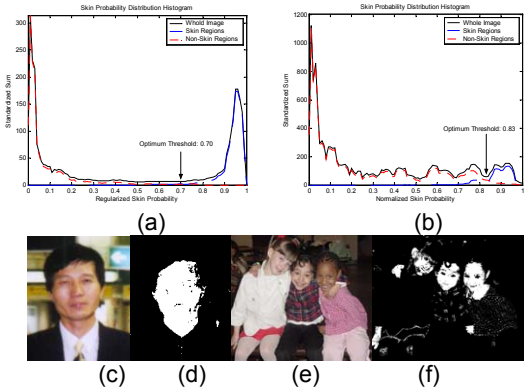

Fig 1. (a)(b) SPDHs of two images, (c)(e) the original images, (d)(f) segmented results by selected optimum thresholds

We adopt machine learning algorithms to learn the rules. For characterizing candidates, we extract 13 features from every candidate, which describes the position, shape and relation with the neighboring valleys of the candidate. Let x_{v_i} and x_{s_i} be the x-coordinate of the i^{th} valley and summit, let h_{v_i} and h_{s_i} be the corresponding heights at the position x_{v_i} and x_{s_i} in the SPDH, we also have $x_{s_{i+1}} < x_{v_i} < x_{s_i}$. For each candidate, the components of feature vector $q_i = [t_1, t_2, t_3, \cdots, t_{13}]$ are given by the following expressions:

Position:

$$t_1 = x_{v_i},\ t_2 = x_{s_i} - x_{v_i},\ t_3 = (x_{s_i} - x_{v_i}) \big/ (x_{s_i} - x_{s_{i+1}})$$

Shape:

$$t_4 = h_{s_{i+1}} \big/ h_{s_i},\ t_5 = h_{v_i} \big/ h_{s_i},\ t_6 = h_{v_i} \big/ h_{\max}$$

$$t_7 = h_{s_i} \big/ h_{\max},\ t_8 = (h_{s_i} - h_{v_i}) \big/ [h_{\max} \cdot (x_{s_i} - x_{v_i})]$$

Area:

$$t_9 = (\int_{x_{v_i}}^{1} h_y \cdot dy) \Big/ (\int_{0}^{1} h_y \cdot dy)$$

Relation with the neighboring candidate:

$$t_{10} = x_{v_i} - x_{v_{i+1}},\ t_{11} = h_{v_{i+1}} \big/ h_{v_i}$$

$$t_{12} = (h_{v_i} - h_{v_{i+1}}) \big/ [h_{\max} \cdot (x_{v_i} - x_{v_{i+1}})],\ t_{13} = h'_{\max} \big/ h_{\max}$$

where $h_{\max}$ is the maximum height of a SPDH, and $h'_{\max}$ is the maximum height in the interval [x_{v_i} , 1].

For comparison, the learning algorithm ANN and C4.5 are both trained. The input is the 13-element feature vector of a certain candidate, and output is a Boolean variable that indicates whether the candidate is optimum. The ANN is designed as a typical two-layers BP network. The first layer has 7 neurons with sigmoid function and the second layer has 1 neuron with binary step function. The feature vector of each candidate in a SPDH is fed to the classifiers respectively by the order of index, until a result "true" outputs. In our experiments, both classifiers are trained and tested on a set of 759 manually labeled images by a ten fold cross-validation techniques. The experimental results show that the ANN is slightly superior to the C4.5, which is listed in the Table2.

Table 2. Comparison of performance between ANN and C4.5

Classification Method	Precision	Recall
ANN	90.8%	92.6%
C4.5	89.8%	90.1%

Since the rules are established automatically and the training set is large enough to include various image contents, our algorithm will find the optimum threshold in most situations. But occasionally they may fail to select the appropriate threshold. For example, the selected threshold is so small that it is unreliable. At this moment a fixed value τ_0 will be selected.

3. Eliminate Confusing Background

Due to the existence of confusing backgrounds, some non-skin regions will be included in above segmented images, even under the optimum threshold. To overcome the problem, we introduce another skin-segmented image by a higher threshold called *upper limit threshold*. Through applying the image relation operation between the segmented images by the optimum and upper limit threshold, the system can eliminate the confusing backgrounds partly.

3.1. Image relation operation

As we know, the detected skin regions in I_i consist two parts: Rss_i – correctly detected regions, and Rns_i – wrongly detected regions. Area(Rxx_i) is the area of regions Rxx_i, I_1 is the segmented image by an optimum threshold, I_2 is the segmented image by an upper limit threshold. Since the upper limit threshold is bigger than the optimum threshold, I_2 has smaller Area(Rss) and Area(Rns) than I_1. We hope that the upper limit threshold can assure that each region of Rss_1 in I_1 will be detected partly in I_2, and Rns_1 will not appear in I_2 at the same time. Then we can apply a relation operation on the two segmented images: the regions in I_1 that overlap with none of regions in I_2 are eliminated as confusing backgrounds, and the remainders in I_1 are kept as real skin regions.

3.2. Finding upper limit threshold

In order to make the above relation operation effective, the selection of upper limit threshold is important. Clue 4 can be employed to help us. Since the summit on the right end of SPDH often overlaps with the summit of the skin hill, the value $[x_{u_1} - (x_r - x_{u_1})]$ may be an appropriate value, where x_{u_1} corresponds to the summit of the hill on the right end of a SPDH, and x_r is on the right hillside, which satisfies $h_r = 0.4 h_{u_1}$. This selected value is situated between the left foot and the summit of the hill, which assures that skin regions can be detected as many as possible and few non-skin regions are included wrongly.

4. Experimental Results

To evaluate the performance of our method, a database with 4000 images is prepared. In which 3000 images come from the ECU database and 1000 images

come from Internet. All the images are manually labeled, 2500(2000+500) images are used for training and 1500(1000+500) images for testing. The training set includes more than 400 million labeled pixels, and covers different races and diverse illuminations.

In our adaptive threshold method, τ_0 equals 0.70 that is used when the selected optimum threshold is unreliable. It has been verified optimum for the fixed threshold method in the training set. The detected results are shown in Table 3, and the results of the fixed threshold method are also listed for comparison. Because no parameters can be adjusted in our method, we can't draw a ROC curve. In order to make the evaluation easier, we choose two special result sets from the ROC curve of the fixed threshold method to compare with ours. One set has an approximate true positive with the results of our method, while the other set has an approximate false positive. From Table 3 we can find that our method outperforms the fixed threshold method in our test set.

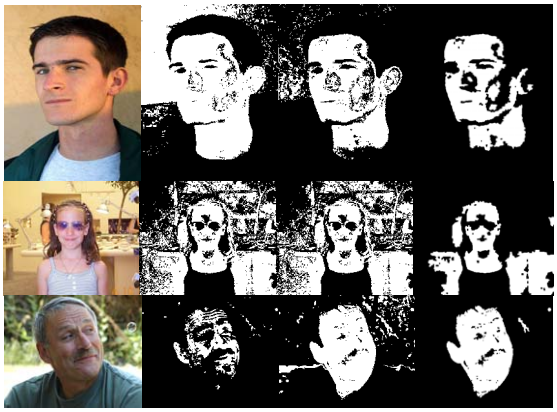

Fig 2. Some experimental results. (a) Original images; (b) results of fixed thresholds; (c) results of adaptive thresholds; (d) final results after eliminating the confusing backgrounds

Some examples are shown in Fig 2. In the original image in the first row, the background has the similar color with the human's face, therefore it is hard to separate from the skin regions. The fixed threshold can't distinguish the tiny difference and regard the confusing backgrounds as skin. In contrary to that, our proposed method can find automatically the adaptive threshold 0.88 which separate the face and background accurately. Another extreme situation is shown in the third row. The light on the man's face changes drastically, about half of the face is in shadow, which will be lost if using the fixed threshold 0.70 to segment. But our method can successfully cope with the situation, the segmented threshold is reduced to 0.23 adaptively and few non-skin regions are included wrongly. From the second row, we will also find that the proposed image relation operation can eliminate the backgrounds effectively when skin regions and confusing backgrounds can't be separated only by thresholds.

Table 3. Comparison result between our method and the fixed threshold method

Method	True positive	False positive
Our method	87.1%	10.3%
Fixed threshold (τ=0.6)	87.1%	15.5%
Fixed threshold (τ=0.8)	78.7%	10.3%

5. Acknowledgments

This work has been financed by the National Hi-Tech R&D Program (the 863 Program) under contract No.2003AA142140.

6. References

[1] Dai, Y., and Nakano, Y., "Face-texture Model-based on SGLD an Its Application in Face Detection in A Color Scene", Pattern Recognition, 1996, pp. 1007-1017

[2] Hsu, R.L., Abdel-Mottaleb, M., and Jain, K., "Face Detection in Color Images", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, 2002, pp. 696–706

[3] Jones, M.M., and Rehg, J.M., "Statistical Color Models with Application to Skin Detection", CVPR, Vol. 2. Fort Collins, Colorado, 1999, pp. 274-280

[4] M. Yang and Narendra Ahuja, "Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Databases", SPIE Proceedings Storage and Retrieval for Image and Video Databases VII, 1999, pp. 458-466

[5] Huynh-Thu, Q., Meguro, M., and Kaneko, M., "Skin-color Extraction in Images with Complex Background and Varying Illumination", IEEE Workshop on Applications of Computer Vision. Orlando, Florida, USA, 2002, pp. 280-285

[6] Phung, S.L., Chai, D., and Bouzerdoum, A., "Adaptive Skin Segmentation in Color Images", IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings, Vol. 3, 2003, pp. 353-356

[7] Cho, K.M., Jang, J.H., and Hong, K.S., "Adaptive Skin-Color Filter", Pattern Recognition, 2001, pp. 1067-1073

[8] Brand, J., and Mason, J.S., "A Comparative Assessment of Three Approaches to Pixel-Level Human Skin-Detection", ICPR, Vol. I, 2000, pp. 1056-1059