ROBUST HUMAN SKIN DETECTION IN COMPLEX ENVIRONMENTS

Ehsan Fazl Ersi, John Zelek
Dept. Of System Design Engineering, University of Waterloo, N2L 3G1 Waterloo, ON, Canada

Keywords: skin detection, image brightness levels, neighborhood information, local entropy thresholding.

Abstract: Skin detection has application in people retrieval, face detection/tracking, hand detection/tracking and more recently on face recognition. However, most of the currently available methods are not robust enough for dealing with some real-world conditions, such as illumination variation and background noises. This paper describes a novel technique for skin detection that is capable of achieving high performance in complex environments with real-world conditions. Three main contributions of our work are: (i) processing each pixel in different brightness levels for handling the problem of illumination variation, (ii) proposing a fast and simple method for incorporating the neighborhood information in processing each pixel, and (iii) presenting a comparative study on thresholding the skin likelihood map, and employing a local entropy technique for binarizing our skin likelihood map. Experiments on a set of real-world images and the comparison with some state-of-the-art methods validate the robustness of our method.

1 INTRODUCTION

In recent years, skin detection has been considered as an active research topic, due to its important role in several applications such as human body part detection and tracking, especially faces (Ming-Hsuan, 2002) and hands (Xiaojin, 2000), human motion analysis (Aggarwal, 1997), and more recently face recognition[1]. Skin Pixels are usually detected using the color information, because (i) it is computationally not expensive, (ii) it is invariant against geometrical transformations (e.g., rotation, scaling and shape changes), and (iii) it can provide a reasonable degree of separability between skin and non-skin classes.

However, color information suffers from sensitivity to illumination variation and besides; pixel-wise color processing does not provide enough information for distinguishing between human skin pixels and background object pixels with skin-like colors. This problem can limit the application of color information in skin detection systems which are used in complex environments.

In this paper we propose an algorithm for accurate detection of human skin, under real-world conditions such as illumination variation and skin-like background colors. There are three main contributions of our skin detection method. We will introduce each of these ideas briefly below and then describe them in details in the subsequent sections.

Processing each pixel in different brightness levels is the first contribution; thus, handling the problem of illumination variation. Each brightness level is generated by increasing or decreasing the color intensities of the original image. Once the brightness levels are generated, the skin likelihood

[1] The identix company claimed that by incorporating skin surface analysis, the accuracy of their facial recognition system, Faceit, improved by at least 20-25%. See: http://www.identix.com/trends/skin.html

Fazl Ersi E. and Zelek J. (2006).
ROBUST HUMAN SKIN DETECTION IN COMPLEX ENVIRONMENTS.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 27-34
DOI: 10.5220/0001376300270034

of each pixel is computed by summing the skin probability[2] of the colors of that pixel in different brightness levels.

The second contribution is a simple and very fast method for incorporating the neighborhood information in processing each pixel. We increase the skin likelihood of each pixel by the average of the likelihood of its direct neighbors. Therefore, the likelihood of a pixel with skin-like neighbors will be increased much more than the likelihood of a pixel with non-skin-like neighbors, although, the skin likelihood of both pixels might be the same. This helps to ignore the small skin like background objects and disregard the small holes in skin regions.

The third major contribution is an evaluation and comparison of the state-of-the-art thresholding methods for binarizing the skin likelihood map. Most of the recently proposed methods for skin detection have used either a constant threshold or some simple histogram-based adaptive thresholding strategy for binarizing their skin likelihood map. However, experimentally we show that the selection of different thresholding strategies can significantly affect the performance of a skin detection method. Using the results of our comparative study, we selected the *local entropy* thresholding method, which is described in details in sub-section 5.2.

Our skin detection system has been trained and tested on the *db-skin* dataset[3] which contains 102 real-world images with changing lighting conditions and/or complex backgrounds (surfaces and objects with skin-like colors). Our method achieved an accuracy rate of 92.1% for a false positive rate of 17.2%, which was superior to the results of other evaluated state-of-the-art methods.

[2] The probability distributions in skin and non-skin classes are obtained from the color histograms of a set of representative images in RGB color space, with skin and non-skin pixels labeled manually.

[3] http://skin.li2.uchile.cl/db1

The remainder of the paper has been organized as follow: Section 2 presents a brief overview on the existing skin detection algorithms; sections 3, 4 and 5 describe different parts of our proposed skin detection method; our experimental results and comparison with other algorithms are illustrated in section 6; and finally, in section 7 some conclusions are given.

2 LITERATURE REVIEW

Based on Martinkauppi's comparative study (Martinkauppi, 2003), the currently available skin detection strategies can be divided into two groups: (i) algorithms which classify a pixel as skin color, if its color is inside some defined region in color space (Dai, 2002) (Hsu, 2002); and (ii) algorithms which classify a pixel as skin color, if its color has a higher than a selected threshold probability. The methods in the second group can be non-parametric like histograms (Jones, 2002), semi-parametric like self organizing maps (Piirainen, 2000) or neural networks (Son, 2001), or parametric assuming a certain distribution, like Gaussian or Gaussian mixtures (Comaniciu, 2000).

The solutions proposed by these approaches for handling the real-world conditions, especially illumination variation, are limited to three: (i) color correctness (e.g., (Hsu, 2002)), (ii) illumination component dropping (e.g., (Stern, 2002)), and (iii) using neighborhood information (e.g., (Ruiz-del-Solar, 2004)). However, color correctness and illumination component dropping have been shown to be not effective in all situations. Funt et al. (Funt, 1998) have shown that the current color correction methods do not necessarily provide better results. Also, it has been shown by Jayaram et al. (Jayaram, 2004) that in most situations, the skin detection performance is significantly better with the presence of an illumination component. Using neighborhood information and region growing, as a more robust solution, improves the performance of pixel-wise approaches in complex environments. However, they can not segment isolated skin regions which are present in bad lighting conditions.

3 PROBABILITY DISTRIBUTION OF SKIN COLORS

The skin and non-skin color models used in our method are built from a set of real-world images in

RGB color space, using a histogram learning technique. We used two 32^3 bin histograms to model skin and non-skin colors. Given skin and non-skin histograms, we can compute the probability that a given color *rgb* belongs to the skin and non-skin classes:

$$p(rgb \mid skin) = \frac{s[rgb]}{T_s}, \quad P(rgb \mid \neg skin) = \frac{n[rgb]}{T_n} \qquad \textbf{(1)}$$

where *s[rgb]* is the pixel count contained in bin *rgb* of the skin histogram, *n[rgb]* is the pixel count contained in bin *rgb* of the non-skin histogram, and T_s and T_n are the total counts contained in the skin and non-skin histograms respectively. Having $p(rgb \mid skin)$ and $p(rgb \mid \neg skin)$, we calculate $p(skin \mid rgb)$ for each color *rgb* using Bayes rule:

$$p(skin \mid rgb) = \frac{p(rgb \mid skin)p(skin)}{p(rgb \mid skin)p(skin) + p(rgb \mid \neg skin)p(\neg skin)} \qquad \textbf{(2)}$$

$p(skin \mid rgb)$ indicates the probability of observing skin, given a color *rgb*. The prior probabilities $p(skin)$ and $p(\neg skin)$ are estimated from the overall number of skin and non-skin samples in the training sets (Jones, 2002).

4 GENERATING THE SKIN LIKELIHOOD MAP

The skin likelihood map is a gray-scale image whose gray values represents the likelihood of the pixel belonging to the skin class. In this section we describe our technique for generating the skin likelihood map.

4.1 Processing Different Brightness Levels for Handling Illumination Variation

In our system, we handle the problem of illumination variation by generating a set of different brightness levels of the image, and assigning skin likelihood to each pixel by processing its colors in different brightness levels. Each brightness level can be generated by raising the color component intensities to the power of γ, where γ is:

$$\gamma = \begin{cases} 1-\beta & \beta > 0 \\ 1/(1+\beta) & \beta \le 0 \end{cases} \qquad \textbf{(3)}$$

The modified image is darker if β is between 0 and 1, and is brighter if β is between -1 and 0.

Once the brightness levels are computed, the skin likelihood of each pixel is obtained by performing the following equation,

$$sl(x,y) = \sum_n p(skin \mid rgb^n_{x,y}) \qquad \textbf{(4)}$$

where $rgb^n_{x,y}$ is the color of pixel *(x,y)* in n^{th} brightness level.

The number of required brightness levels can be automatically determined for each image by analyzing its color histogram. However, for the sake of simplicity and fast performance, we only process two brightness levels besides the original image: (i) a darker level with β=*-0.5* and (ii) a brighter level with β=*+0.5*. (see Figure1.a)

4.2 Incorporating Neighborhood Information

Besides the color information, which exhibits a reasonable degree of separability between skin and non-skin classes, there is another source of information that can be used in skin detection: skin regions have low texture and a homogeneous local color distribution. There have been some reports in the literature that incorporate this property in their detection process. However, most of them were either computationally expensive or applicable to only some particular situations (Alibiol, 2001) (Martinkauppi, 2002).

In this paper, we employ a very simple and fast technique for considering the neighborhood information in classifying a pixel into skin or non-skin classes. Assuming a *wxw* window centered at a given pixel *(x,y)*, the skin likelihood of pixel *(x,y)* is increased using the following equation:

$$sl^*(x,y) = sl(x,y) + \frac{1}{w^2}\sum_{i=-\lfloor w/2\rfloor}^{\lfloor w/2\rfloor}\sum_{j=-\lfloor w/2\rfloor}^{\lfloor w/2\rfloor} sl(x+i, y+j) \qquad \textbf{(5)}$$

w can be any odd number between *3* and *min(W,H)*, where *W* and *H* are the width and the heights of the image, respectively.

Although equation 5 is nothing but a set of addition and a division operations, for *w>5* its computational cost can be considerable. Therefore we employ an intermediate representation for *sl*, called the integral image (Viola, 2001), which provides a very fast scheme for computing *sl**.

(a)

(b)

(c)

Figure1: (a) Shows three brightness levels of a sample image, as well as the skin probability of the color of each pixel in different brightness levels, for *β=0, β=-0.5,* and *β=+0.5,* from left to right. (b) Shows the skin likelihood, *sl*, obtained by using Eq.4, and (c) shows *sl**, our final skin likelihood map. As can be seen, *sl**, appears qualitatively more accurate than *sl* and any of the skin probability maps generated in different brightness levels.

The integral image at location *(x,y)* contains the sum of the pixels above and to the left of *(x,y)*. Using the integral image any rectangular sum can be computed in four array references with four simple addition and subtraction operations. (for details on generating the integral image, see (Viola, 2001)).

Figure1.c shows *sl** for the skin likelihood (figure 1.b) of a sample image. As can be seen, the likelihood of a pixel with skin-like neighbors will be increased much more than the likelihood of a pixel with non-skin-like neighbors, although the skin likelihood of both pixels might be the same. This helps to ignore the small skin like background objects and disregard the small holes in skin regions.

After incorporating neighborhood information, we scale the skin likelihood map to 256 gray levels, after which, thresholding is applied.

5 THRESHOLDING THE SKIN LIKELIHOOD MAP

5.1 Thresholding Strategies

In order to segment the skin regions from a skin likelihood map, a thresholding process should be used. Most of the currently available skin detection strategies have not paid sufficient attention to the thresholding process and simply select a fixed value, obtained by analyzing the receiver operating characteristic (ROC) curve. However, we believe that different images of different people with different skins cannot be binarized using a unique fixed threshold value. In this section we compare several state-of-the-art thresholding methods, as well as a number of constant thresholds, to examine if applying an adaptive thresholding strategy can improve the detection performance.

Five methods are used in the skin likelihood thresholding for comparison: Otsu's method (Otsu, 1979), Pal & Pal's local entropy (LE), global entropy (GE), and joint entropy (JE) methods (Pal and Pal, 1989), and Jones's constant thresholding (Jones, 2002). The skin likelihood maps of 27 randomly selected images from the *db-skin* dataset were used for the experiments. Table1 shows the performance of the five adaptive thresholding strategies as well as the performance of three fixed threshold values. Also the performance of three adaptive techniques as well as Jones fixed strategy on a difficult sample image is illustrated in figure 2.

Table1: The performance of different thresholding strategies.

Methods	False Negative	False Positive
Otsu	6.9%	20.7%
Pal & Pal (GE)	8.4%	17.6%
Pal & Pal (LE)	**9.9%**	**16.0%**
Pal & Pal (JE)	8.2%	18.8%
Jones: 128	9.9%	17.7%
Fixed: 110	7.0%	22.2%
Fixed: 150	17.0%	11.4%
Fixed: 200	20.1%	10.2%

Based on our experimental results, we selected Pal & Pal's local entropy technique for thresholding our skin likelihood maps, due to its trade-off between false positive and false negative rates.

5.2 Local Entropy Thresholding Method

Entropy is the measure of the information content in a probability distribution. To provide the probability distribution needed for the entropy measures, a co-occurrence matrix is generated from the input image. It is a mapping of the pixel to pixel grey scale transitions in the image between the neighboring pixel to the right and the pixel below each pixel in the image. From the co-occurrence matrix comes the distribution of grey scale transitions. The candidate threshold divides the co-occurrence matrix into four regions representing within object, within background, object to background, and background to object class transitions (see figure2). Then, the second-order local entropy is computed by using the local entropies of backgrounds and objects:

$$H_{local}^{(2)}(t) = H_A^{(2)}(t) + H_C^{(2)}(t), \qquad \textbf{(6)}$$
$$= -\frac{1}{2}\sum_{i=0}^{t}\sum_{j=0}^{t} p_{i,j}^{A} \log p_{i,j}^{A}$$
$$-\frac{1}{2}\sum_{i=t+1}^{255}\sum_{j=t+1}^{255} p_{i,j}^{C} \log p_{i,j}^{C}$$

In the above equation $H_A^{(2)}$ and $H_C^{(2)}$ are the local entropies of background and objects, respectively. The optimal threshold is found by maximizing the $H_{local}^{(2)}(t)$. For more details about the algorithm see (Pal and Pal, 1989).

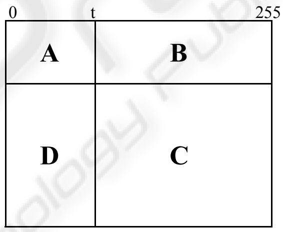

Figure2: Quadrants of a co-occurance matrix. A and C are background and object respectively.

5.3 Post-processing

Once the thresholding process has been done, we perform a simple region growing technique, iteratively, in order to add those pixels which are (i) very close to a skin region, and (ii) have skin likelihoods just under the applied threshold. This helps to cover the skin edges which might be removed during the incorporation of neighborhood information.

6 EXPERIMENTS

Our experiments have been done on the *db-skin* dataset, which contains 102 images under real-world conditions, obtained from Internet and from digitized news videos.

The majority of the images are difficult to segment, due to either bad lighting conditions or complex backgrounds containing surface and objects with skin like colors.


Original image


Skin likelihood map


Jones performance


Otsu performance


Local entropy performance


Global entropy

Figure2: The performances of different thresholding strategies on a sample image with a great amount of skin-like background colors.

The result of applying some state-of-the-art algorithms on a set of 27 selected images from the *db-skin* dataset are reported in (Ruiz-del-Solar, 2004, FGR) and (Ruiz-Del-Solar, 2004, ICIP). The evaluated methods were: *Jones1*, which corresponds to the MoG classifier proposed in (Jones, 2002) using skin color model and a fixed threshold; *Jones2*, the same as Jones1 but with employing non-skin color model as well; *SkinDiff*, which corresponds to the skin detection method proposed in (Ruiz-Del-Solar, 2004, ICIP) (RGB, MoG, and diffusion algorithm); and *HSU*, which corresponds to the skin detection algorithm proposed in (Hsu, 2002) but without the use of whitening compensation.

In order to compare our results with the evaluated methods, we randomly selected 27 images for testing and the rest for training. Since we did not know which images had been used by Solar et al. (Ruiz-del-Solar, 2004, FGR) (Ruiz-Del-Solar, 2004, ICIP) in their experiments, we performed our experiments three times and for each time with different sets of training and testing images. Then we averaged the results. Table2 shows the performance of or skin detection method, in comparison with four previously evaluated methods.

Table2: The performance of our skin detection algorithm in comparison with four state-of-the-art methods at a false positive rate of ~0.17. Our results are the average of three performances in different sets of randomly selected images.

Methods	Detection rate	False positive
HSU	73.0%	17%
Jones1	85.0%	17%
Jones2	86.5%	17%
SkinDiff	88.1%	17%
Our method	**92.1%**	**17.2%**

Even though a considerable amount of processing is employed for implementing our method, a reasonable high processing speed is achieved. It is worth to mention that the cost of processing different brightness levels implemented by LUT (look-up table), or incorporating neighborhood information implemented using integral image representation, was even less than the transformation of the RGB color space to some other color spaces like CIELAB or HIS, which have been used by various researchers. However, the cost of employing an adaptive thresholding strategy, especially the entropic ones, is remarkable. The computational time of our method for processing a 320x240 image is approximately 1.1s on a 3.06 GHz CPU. However, this time is decreased to 0.28s by employing a fixed value threshold.

7 CONCLUSION AND FUTURE WORK

In this paper a novel skin detection algorithm was proposed for handling the real-world situations, such as bad lighting conditions or skin-like background colors. Three contributions of our work are: (i) processing each pixel in different brightness levels for handling the problem of illumination variation; (ii) presenting a fast and simple method for incorporating the neighborhood information in processing each pixel; and (iii) presenting a comparative study on thresholding the skin likelihood map, and employing local entropy technique for binarizing our skin likelihood map. The details of our method are described and the detection performance is compared with some state-of-the-art methods using a set of real-world images, obtaining better results.

One of the directions that we are considering for future work is to incorporate texture and shape into our skin detection method. Furthermore, we intend to apply our skin detection strategy to additional applications such as nudity detection and adult image filtering.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Communication and Information Technology Ontario (CITO) for partially supporting this work.

REFERENCES

Aggarwal, J.K.; Cai, Q., 1997. Human motion analysis: a review. In Nonrigid and Articulated Motion Workshop 1997, IEEE Proceedings.

Alibiol, A. and Torres, L., 2001. Unsupervised color image segmentation algorithm for face detection applications. In Proc. 3rd IEEE International Conference on Image Processing.

Comaniciu, D. and Ramesh, V., 2000. Robust detection and tracking of human faces with an active camera. In Proc. 3rd IEEE International Workshop on Visual Surveillance.

Dai, Y. and Nakano, Y., 2002. Face-texture model-based on SGLD and its application in face detection in a color scene. Pattern Recognition, 29(6).

Funt, B. and Barnard, K., 1998. Is machine color constancy good enough?. In Proc. 5th European Conference on Computer Vision

Hsu, R. L., Abdel-Mottaleb, M. and Jain AK., A. K.,2002. Face detection in color images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5).

Jayaram, S.; Schmugge, S.; Shin, M.C.; Tsap, L.V., 2004. Effect of colorspace transformation, the illuminance component, and color modeling on skin detection. In CVPR 2004, Proceedings of the 2004 IEEE Computer Society Conference Volume 2.

Jones, M.J. and Rehg, J.M., 2002. Statistical color models with application to skin detection. International Journal of Computer Vision, 46(1).

Pal, N. and Pal, S., 1989. Entropic thresholding. In Sygnal Process.

Piirainen, T., Silv´en, O. and Tuulos, V., 2000. Layered selforganizing maps based video content classification. Workshop on Real-time Image Sequence Analysis.

Martinkauppi, B., 2002. Face color under varying illumination – analysis and applications. Ph.D. thesis, University of Oulu.

Martinkauppi, B.; Soriano, M.; Pietikainen, M., 2003. Detection of skin color under changing illumination: a comparative study. In Image Analysis and Processing, 2003.Proceedings. 12th International Conference.

Ming-Hsuan, Y.; Kriegman, D.J.; Ahuja, N., 2002. Detecting faces in images: a survey.

In Pattern Analysis and Machine Intelligence, IEEE Transactions, vol 24.

Otsu, N., 1979. A threshold selection method from gray-level histograms. In IEEE Trans. Syst. Man Cybern.

Ruiz-Del-Solar, J. and Verschae, R., 2004. Robust skin segmentation using neighborhood information. In. Proc.IEEE International Conference on Image Processing (ICIP).

Ruiz-del-Solar, J. and Verschae, R., 2004. Skin detection using neighborhood information. In Automatic Face and Gesture Recognition, 2004 (FGR). Proceedings. Sixth IEEE International Conference.

Son, L.M., Chai, D. and Bouzerdoum, A., 2001. A universal and robust human skin color model using neural networks. Proc. IJCNN '01 International Joint Conference on Neural Networks, vol. 4.

Stern, H; Efros, B., 2002. Adaptive color space switching for face tracking in multi-colored lighting environments. In Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference.

Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In Proc. IEEE CVPR.

Xiaojin, Z.; Jie Y.; Waibel, A., 2000. Segmenting hands of arbitrary color. In Automatic Face and Gesture Recognition Proceedings. Fourth IEEE International Conference.