

ROBUST SKIN SEGMENTATION USING NEIGHBORHOOD INFORMATION

Javier Ruiz-del-Solar and Rodrigo Verschaë

Dept. of Electrical Engineering, Universidad de Chile, Santiago, CHILE

E-mail: jruizd@ing.uchile.cl, rverscha@ing.uchile.cl

ABSTRACT

Skin detection or segmentation is employed in many tasks related to the detection and tracking of humans and human-body parts. However, skin detection is not robust enough for dealing with some real-world conditions, such as changing lighting conditions and complex backgrounds containing surfaces and objects with skin-like colors. This situation can be improved by incorporating context information in the skin detection process. For this reason in this article a skin detection approach that uses neighborhood information is proposed. This idea is implemented through a diffusion process that allows a robust segmentation of skin regions at a high processing speed.

1. INTRODUCTION

Skin detection or segmentation is a very popular and useful technique for detecting and tracking human-body parts, specially faces and hands. Its most attractive properties are: (i) high processing speed due to its low-level processing, and (ii) invariance against rotations, partial occlusions and pose changes. However, standard skin detection techniques are not robust enough for dealing with complex environments. Changing lighting conditions and complex backgrounds containing surfaces and objects with skin-like colors are major problems, limiting its use in practical real-world applications. For solving the mentioned drawbacks, many groups have centered their research on selecting the color-space most suitable for skin detection. Many different color models have been employed, among them: RGB, normalized RGB, HIS-HSV, YCbCr, YIQ, YES, YUV, CIE XYZ, CIE LUV, and ab (see some references in [5]). We believe that just selecting the “best” color space does not solve the mentioned problem and that only marginal improvements can be achieved (see for example [6][2]).

Some authors have centered their efforts on modeling the physical properties of the acquisition process. In [5] the physical properties and the calibration of the camera to be employed in the acquisition process are modeled, while in [3] a color constancy method (white balancing) for obtaining

some invariance against changing illumination is employed. We believe that these approaches are not general enough and they can be implemented only in some special cases.

Other authors have used statistical models for solving the skin/non-skin classification problem. Most successful approaches are Mixture of Gaussians (MoG) [7] and histogram models [4]. They differ in the parametric or non-parametric form of computing the skin/non-skin probabilities. (Although is a popular belief that histogram models have a high processing speed because they employ LUTs (look-up-tables), in this article we show that the same procedure can be used for the MoG). We think that statistical models are in the right direction for dealing with the real-world problems of skin detection, however they miss the benefits of using contextual information.

All mentioned approaches are based on the same *pixel-wise* processing paradigm, in which each image pixel is individually analyzed. We think that this paradigm should be extended; context information should be incorporated in the skin detection process. Human beings can detect skin in real scenes, or in pictures and videos without problems. However, for a human being the classification of a single pixel as skin or non-skin is a very difficult task. Humans use many high-level processes to assist the detection of skin (detection of hair, detection of clothes, etc), and also some simple diffusion mechanisms, employed in general, for the segmentation of colors and textures. Taking this fact into consideration we propose the use of context information in the skin detection process. Spatial context can be employed for skin detection in still images, while time context can also be employed when detecting skin in videos. When dealing with high-level tasks such as face and person detection, high-level information (e.g. detected faces) can assist the skin detection process.

Following this idea in this article is proposed a robust skin segmentation approach that uses neighborhood information, i.e. the decision about the class (skin or non-skin) of a given pixel considers information about the pixel’s neighbors. Although this idea has been employed before in the post-processing of skin-detections (e.g. holes in skin detected areas are

filled using morphological dilation and isolated skin pixels are deleted using morphological erosion), and more recently in [1] using unsupervised segmentation and region grouping, in this work a different and more radical approach is employed: a diffusion process is implemented for determining the skin pixels. The aim of this process is not just the grouping of neighbor skin pixels, but also the determination of skin areas where the “skiness” of pixels are larger than a minimal threshold and where the “skiness” of neighbor pixels changes smoothly. This process is controlled using three thresholds: one for determining the minimal acceptable “skiness” of a skin pixel (T_{min}), one for controlling smooth changes of the pixel “skiness” between neighbors (T_{diff}), and a third one for determining the seeds of the diffusion process (T_{seed}). Before diffusion the “skiness” measure, which can either correspond to the probability of being a skin pixel when a statistical model is used, or to a membership degree when a fuzzy approach is employed, should be calculated. Even though a diffusion process is employed for implementing the skin segmentation, a high processing speed is achieved (see section 3).

2. PROPOSED SKIN DETECTION APPROACH

For the sake of simplicity in the description of our algorithms we will use the spatial-range domain as image space. In this domain each image pixel has two parts, a spatial part (position) and a range part (intensity), where the range part may be written as a function of the spatial part:

$$x_j = (x_j^s, x_j^r) = (x_j^s, I(x_j^s)) \quad (1)$$

The superscripts s and r denote the spatial and range parts of the pixels. The spatial-range domain has a dimension of $d=r+2$, with $r=1$ for grayscale images and $r=3$ for color images. The “2” value represents the bi-dimensional spatial characteristic of images.

2.1. Pixel-Wise Classification

The “skiness”, i.e. the belonging of a pixel to the skin class, is calculated as a real number between 0 and 1. When a statistical model is used, this number corresponds to the probability of being a skin pixel, while when a fuzzy approach is employed, it corresponds to a fuzzy membership degree. In this last case any crisp pixel-wise classification algorithm can be employed as a basis algorithm, and its results can be then fuzzified. For instance, if an elliptical skin cluster is used as pixel-wise classification method (on a given color space), before classification the cluster is fuzzified: a value 1 is assigned to the center of the cluster, and decreasing values are assigned to the other colors.

When a probabilistic model is chosen, the best alternatives are MoG [7] and histogram models [4], being their main difference the way in which the skin/non-skin probabilities are calculated. MoG is a parametric model and thereafter less training samples are required for obtaining good probabilities’ estimators. Usually it is argued that histograms are faster because they are implemented using LUTs [4], but LUTs can also be used for implementing the MoG (see section 2.3). Therefore, a similar performance can be achieved using any of these models. Although in [4] slightly better results are obtained using histograms than using MoG, the reason seems to be the use of a non-skin model in the former case.

A general pixel-wise classification algorithm is presented in figure 1. As it can be seen, this algorithm imposes no conditions on the color space, or on the statistical model or fuzzification function being used.

```

pixel_wise_classification( $I, I_{skin}$ ) {
  foreach  $x_i \in I, y_i \in I_{skin}$ 
     $y_i^r = g_{skin}(x_i^r)$ 
}
 $I$ : input image;  $I_{skin}$ : image with skin probabilities or membership degrees as pixel values.
 $g_{skin}$ : probabilistic model or fuzzification function.

```

Figure 1. Pixel-wise classification algorithm.

2.2. Spatial Diffusion

The final decision about the pixel’s class is taken using a spatial diffusion process that takes into account context information. In this process a given pixel will belong to the skin class if and only if its distance, in this case the Euclidean distance, calculated in a given color space, with a direct diffusion-neighbor that already belongs to the skin class, is smaller than a certain threshold (T_{diff}). The seeds of the diffusion process are pixels with a high probability of being skin or with a large membership to the skin class, i.e. their probability or membership degree is larger than a given threshold (T_{seed}). The extension of the diffusion process is controlled using a third threshold (T_{min}), which defines the minimal probability or membership degree allowed for a skin pixel. This diffusion algorithm is shown in figure 2.

2.3. Computational Implementation

The ideas described in the former two sections are implemented by the *SkinDiff* algorithm. *SkinDiff* uses a RGB color space, the MoG model implemented using LUTs and the diffusion algorithm presented in section 2.2. By looking at the pixel-wise classification and diffusion algorithms it can be noted that both can be joined in a single processing step. If this is done, it is also not necessary to store the skin probabilities in the LUT, but only the information concerning the

following three situations: skin probability larger than T_{seed} , smaller than T_{min} or in $[T_{seed}, T_{min}]$. Therefore for each possible RGB combination, only 2 bits needs to be stored: LUT[\cdot].seed (1 means $\geq T_{seed}$) and LUT[\cdot].min (1 means $\geq T_{min}$). For an adequate implementation of the LUT, the colors in each channel are quantized. The diffusion algorithm is further optimized using an implicit stack instead of recursion. A last optimization is the use of the square of the Euclidean distance. The optimized diffusion algorithm using LUTs and the implicit stack is shown in figure 3.

```

diffusion_algorithm( $I_{skin}, S_{skin}$ ) {
  find_seeds( $I_{skin}, S_{seed}$ )
  foreach  $s \in S_{seed}$ 
    PushItem( $S_{skin}, s$ )
    diffusion_rec( $s, S_{skin}$ )
}
find_seeds( $I_{skin}, S_{seed}$ ) {
  foreach  $x_i \in I_{skin}$ 
    if  $x_i^r > T_{seed}$ 
      PushItem( $S_{seed}, x_i$ )
}
diffusion_rec( $s, S_{skin}$ ) {
  if  $s^r \geq T_{min}$ 
    foreach  $x_j^s \in ConnectedNeighbors(s^s)$ 
      if  $x_j \notin S_{skin}$ 
        if  $|x_j^r - s^r| < T_{diff}$ 
          PushItem( $S_{skin}, x_j$ )
          diffusion_rec( $x_j, S_{skin}$ )
}

```

I_{skin} : image with skin probabilities or membership degrees as pixel values. S_{skin} : final set of skin pixels, the output of the algorithm. S_{seed} : set of seed pixels.
 PushItem(S, x) adds x to the end of S .
 ConnectedNeighbors(s) returns the 8-connected neighbors of s in the 2D lattice of the image.

Figure 2. Diffusion algorithm.

3. RESULTS

For testing the performance of the proposed *SkinDiff* algorithm, in real-world conditions, we selected a set of 27 images obtained from Internet of from digitized news videos from ABC and CNN. The selected images are considered difficult to segment, because they have either changing lighting conditions or complex backgrounds containing surfaces and objects with skin-like colors. The selected images, as well as, their ground truth information will be made available for future studies in a near future.

For testing the performance of *SkinDiff*, 3 baseline algorithms were employed: **Jones1**, **Jones2** and **HSU**. Jones1 corresponds to the MoG classifier proposed in [4], using only the skin color and a fixed decision

threshold (T_{fix1}). Jones2 corresponds to the MoG classifier proposed in [4], using the skin color and the non-skin color model, and a fixed decision threshold (T_{fix2}). Jones 1 and Jones 2 were implemented using the same parameters employed in the original work. These parameters were also employed for the MoG classifier of *SkinDiff*. HSU corresponds to the skin detection algorithm proposed in [3] (YCbCr, elliptical cluster model), but without use of whitening compensation, for fairness in the comparison with the other algorithms. In all methods, no post-processing (e.g. opening / closing) was employed.

```

diffusion_algorithm( $I, S_{skin}$ ) {
  find_seeds( $I, S_{seed}$ )
  foreach  $\tilde{s} \in S_{seed}$ 
    PushItem( $S_{skin}, \tilde{s}$ )
    PushItem( $Stack, \tilde{s}$ )
    while( $Stack \neq \phi$ )
       $s \leftarrow PopItem(Stack)$ 
      if LUT[ $s^r$ ].min
        foreach  $x_j^s \in ConnectedNeighbors(s^s)$ 
          if  $x_j \notin S_{skin}$ 
            if  $|x_j^r - s^r|^2 < T_{diff}^2$ 
              PushItem( $S_{skin}, x_j$ )
              PushItem( $Stack, x_j$ )
}
find_seeds( $I, S_{seed}$ ) {
  foreach  $x_i \in I$ 
    if LUT[ $x_i^r$ ].seed
      PushItem( $S_{seed}, x_i$ )
}

```

I : input image. S_{skin} : final set of skin pixels, S_{seed} : set of seed pixels. $Stack$: stack data structure.

Figure 3. Diffusion algorithm using LUTs and a stack.

In figure 4 are shown ROC curves of the 4 compared algorithms. For *SkinDiff* the curves were built using different values of T_{min} and T_{diff} , and also using LUTs with different number of quantized colors (e.g. LUT-64 means 64 colors in each color channel). For Jones1/Jones 2 the parameters to be changed were T_{fix1}/T_{fix2} , while for HSU it was the ellipse radii, which were change proportionally.

The ROC curves of *SkinDiff* were obtained as follows: for a given value of T_{min} and T_{diff} , T_{seed} was moved between T_{min} and 1. By doing that many curves can be obtained, but in figure 4 only some of them are shown. By analyzing the ROC curves it can be noticed that *SkinDiff* implementations outperform the baseline algorithms. In table 1 are also shown the processing time (PT) of the different algorithms for some points of the ROC curves (27 images). These points correspond to similar detection rates (DR), but very different false positive rates (FPR). It can be observed

that SkinDiff LUT-128 and SkinDiff LUT-64 achieve the best compromise between PT, DR and FPR.

Some features of SkinDiff that will be detailed analyzed in a future publication: (i) It performs very well when detecting skin on bright image areas; (ii) Skin seeds seldom appear in background areas, which allows the adequate processing of backgrounds; and

(iii) less post-processing of the skin detections is required, it is not necessary to filter isolated skin pixels and the filling of holes can be done using a smaller dilation filter (faster application). As a future work authors want to extend the algorithm to videos by using temporal context.

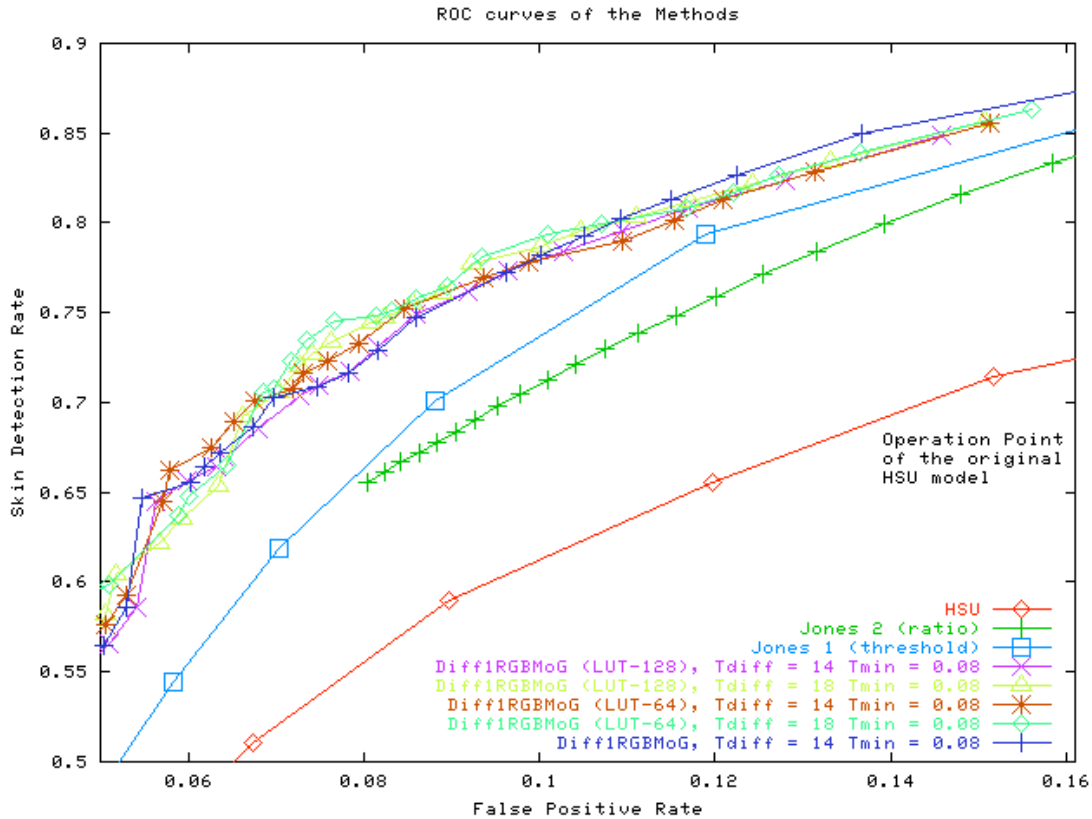


Figure 4. ROC Curves of the four compared algorithms (see main text for details).

Table 1. Total processing time (PT) for the 27 images.

Algorithm	Total PT [s]	Detection Rate	False Positive Rate
HSU	9	0.714	0.152
Jones1 LUT-128	8	0.701	0.088
Jones2 LUT-128	8	0.705	0.098
SkinDiff	32	0.722	0.072
SkinDiff LUT-128	11	0.709	0.074
SkinDiff LUT-64	10	0.709	0.070

Acknowledgements

This project was partially funded by the FONDECYT project 1030500 (Chile) and by the join CONICYT-BMBF project “Human-Machine Affective and Collaborative Interaction via Multimodal Interfaces”.

5. REFERENCES

[1] A. Albiol, L. Torres, Ed. Delp, “An Unsupervised Color Image Segmentation Algorithm for Face Detection Applications”, *IEEE Int. Conf. on Image Proc. – ICIP 2001*, Greece, 2001.

[2] A. Albiol, L. Torres, Ed. Delp, “Optimum Color Spaces for Skin Detection”, *IEEE Int. Conf. on Image Proc. – ICIP 2001*, Greece, 2001.

[3] R. L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, “Face detection in color images”, *IEEE Trans. on Pattern Anal. and Machine Intell.* 24(5): 696-706, 2002.

[4] M.J. Jones, and J.M. Rehg, “Statistical color models with application to skin detection”, *Int. Journal of Computer Vision* 46(1): 81-96, 2002.

[5] B. Martinkauppi, *Face Color under Varying Illumination – Analysis and Applications*, Doctoral Thesis, University of Oulu, Finland, 2002.

[6] M. Shin, K. Chang, and L. Tsap, “Does colorspace transformation make any difference on skin detection?”, *Proc. IEEE Workshop on Appl. of Computer Vision 2002*, Florida, USA, 2002.

[7] M. H. Yang, and N. Ahuja, “Detecting human faces in color images”, *Proc. IEEE Int. Conf. on Image Processing*, Chicago, Illinois, USA, 1: 127-130, 1998.