Colour based human skin segmentation using graph-cuts

L. Lattari, A. Conci, A. Montenegro, E. Clua
Instituto de Computação
Universidade Federal Fluminense
Niterói, RJ, Brazil
{llattari, aconci, anselmo, esteban}@ic.uff.br

V. F. Mota, M. B. Vieira
Departamento de Ciência da Computação
Universidade Federal de Juiz de Fora
Juiz de Fora, MG, Brazil
{virginia.fernandes, marcelo.bernardes}@ice.ufjf.br

Abstract— This paper proposes a new method to deal with the problem of automatic human skin segmentation on RGB color space model. The problem is modeled as a min-cut problem on a graph whose vertices represent the image colour characteristics and the edges represent the costs of labeling the vertices as skin or non-skin, based on an energy function defined in terms of a database of skin tones. Our method was evaluated under several circumstances, indicating when correct or incorrect results were generated. The overall experiments have shown that our automatic method is simple, efficient and is competitive with the existing methods in the literature.

Keywords- skin segmentation; pixel-based classification ; colour database; colour based segmentation; energy function; graph-cuts.

I. INTRODUCTION

Human skin segmentation is a research field with many applications such as video surveillance, face or hand gesture recognition, content-based visual information retrieval (CBVIR), filtering on the web and others. This is a fundamental task for any application that looks for human sequences on image and video streams. In many situations, very little human intervention must occur (as when the data to be processed is represented by a huge database). This paper presents a new automatic per-pixel human skin segmentation method, which is highly customizable and yields good results compared with other relevant works.

Many methods and models have been proposed to segment human skin. Hu et al [1] obtained skin regions after detection due by a single Gaussian Model and a Gaussian Mixture Model, respectively for skin and non-skin. After that, a noise suppression, a segmentation and a filtering stage are made, obtaining final result. The segmentation method is the same as us, using graph-cuts. Their results are very good, even with illumination variation. Sigal et al [2] considered time-varying illumination with multiple sources and colours, using a Markov model to predict evolution of skin histogram on HSV colour space over time, obtaining good results. Jedynak et al [3] compared three approaches for skin detection and different from most works which gives a float number probability solution instead of binary. Final results are measured by a ROC Curve and thresholded by user, which means that is not automatic. Also, it involves Markov and Parameter Estimation methods, giving good results.

Most skin segmentation methods are colour based approaches and introduce a colour metric which considers the distance of the pixel to a specific colour [1,2,5,6]. It is not a trivial problem to solve, since objects and backgrounds exist in a large variety of colours, including skin tones. This paper is based on the premise that skin colours form a small and unique subset of the RGB colour space, which makes it easier to solve this specific case of segmentation [2]. In this approach the segmentation is modeled as a min-cut problem on a direct graph [6,7] and each edge has a well defined cost. Every node corresponds to a pixel and each edge weight is obtained by heuristics based on pixel colours. The result of the optimization process, this method separates the graph into two disjoint pixel sets: skin and non-skin regions. One of the advantages of our method is that it is very simple to implement. Moreover, it yields good results when compared to others. It is suitable for a parallel implementation as, for example, a GPU computing approach.

The rest of the paper is organized as follows. The next section summarizes the graph-cut method which is the basis of the proposed method. Section III describes how the database is defined and how it is used in our model. The methodology is explained in Section IV. Our results are presented in Section V. Finally, in Section VI, conclusions and future works are outlined.

II. SEGMENTATION BY GRAPH-CUTS

A. Energy Function

It is possible to find a characteristic function of an object defined in a given domain by minimizing an objective function, i.e., given a set M, we have to find the characteristic function X which is the minimum argument of a function [6] and partition sets. A widely used objective function in image segmentation is the energy of Gibbs defined as ([6, 7, 8])

$$E(X) = \sum_{x_i \in V} E_1\big(X(x_i)\big) + \lambda \sum_{x_i, x_j \in \varepsilon} E_2\big(X(x_i), X(x_j)\big), \qquad (1)$$

where x_i and x_j are elements of the set to be segmented, V is the set of elements, ε is the set of connected

This work was performed under FAPERJ, FAPEMIG and CAPES grants.

elements and λ is a weight. E_1 is the term that defines a cost for each x_i to belong to one of the sets. Aiming to minimize the objective function, this cost should be inversely proportional to the probability of x_i to belong to the set. It is usually given as

$$\begin{array}{lll} E_1(X(x_i)=1)=0 & E_1(X(x_i)=0)=\infty & \forall x_i \in O \\ E_1(X(x_i)=1)=\infty & E_1(X(x_i)=0)=0 & \forall x_i \in B \\ E_1(X(x_i)=1)=\phi(\rho_o) & E_1(X(x_i)=0)=\phi(\rho_b) & \forall x_i \in N \end{array} \quad (2)$$

where O is the set of object elements x_i, B is the set of background elements, N is the set of pixels whose labels are unknown and ϕ is a function inversely proportional to its parameters terms ρ_o and ρ_b.

E_2 is a term that defines a penalty for labeling two connected elements with different labels. This penalty depends on the similarity of both elements: very similar elements have high probability of belonging to the same set. In this case the resulting cost must be high; otherwise it has a small value.

B. *Graph-cuts*

The energy function in (1) is used for defining capacities $w \geq 0$ for each edge $(u,v) \in E$ on a directed graph $G=(V,E)$[9]. Two terminal nodes are defined: a source *s* and a sink *t*, corresponding to the two labels that can be assigned to pixels. Each non-terminal node in the graph will have an edge connected with *s* and *t*. We assume that every vertex lies on some path from the source to the sink.

Two types of edges are then defined: T-links and N-links. T-links connects pixels to terminals. N-links connects pairs of neighbor pixels. Their costs are based in the terms E_1 and E_2, respectively.

A cut is a partition subset of *V* into two subsets S and T, that is, V = S + T such that $s \in S$ and $t \in T$. Considering the cost of a cut C as the sum of costs, our objective is to find a cut that has the minimum cost among all cuts. Ford and Fulkerson [10] have shown that the maximum flow *f* of *G* is equivalent to the cost of a minimum cut. That is important because it is more efficient to compute a maximum flow than a minimum cost cut.

III. Human Skin Database

In order to define an energy function for skin detection, we propose the use of *n* images where skin regions and non-skin regions are explicitly marked by some user. The idea is to use these multiple images to infer the likelihood of a pixel colour to represent a skin or non-skin region.

In each image: I_i, $0 < i \leq n$, of the database, n_o pixels are marked as skin, forming an object region $O \subset I_i$, and n_b pixels are marked as background, forming a region $B \subset I_i$. Note that $O \cup B$ is not necessarily a partition of I_i. This means that the user does not have to label all pixels of the image.

The database is then defined by *n* sets, each one composed by k_o colours of O_i and k_b colours of B_i. These k_o and k_b colours are computed using a quantization method capable of finding the best colours to represent each gamut sub region. In our work, we use the K-means method [11] for simplicity.

IV. Proposed Method

Let $CO_{m,i}$ be the *m*-th mean object colour and $CB_{l,i}$ be the *l*-th mean background colour of the *i*-th image of the database, where $0 < m \leq k_o$, $0 < l \leq k_b$ and $0 < i \leq n$. It is possible to define the similarities by the following functions

$$\begin{aligned} d_{o,i} &= \min_m \left\| C_p - CO_{m,i} \right\| \\ d_{b,i} &= \min_l \left\| C_p - CB_{l,i} \right\| \end{aligned} \quad (3)$$

where C_p is the colour of the pixel *p* that we need to label as skin or not, $d_{o,i}$ and $d_{b,i}$ are the minimum distance between C_p and the colours of the *i*-th object set and the colours of the *i*-th background set in the database, respectively.

Similar to [7], our energy terms for (1) are then defined as

$$\phi(\rho_o) = E_1(X(p)=1) = \frac{\sum_{i=1}^{n} \frac{d_{o,i}}{d_{o,i}+d_{b,i}}}{n} \quad (4)$$

$$\phi(\rho_b) = E_1(X(p)=0) = \frac{\sum_{i=1}^{n} \frac{d_{b,i}}{d_{o,i}+d_{b,i}}}{n} \quad (5)$$

$$E_2(X(p),X(q)) = \frac{|X(p)-X(q)|}{\left\| C_p - C_q \right\|^2 + 1} \quad (6)$$

$$\lambda = 1$$

where C_p and C_q are the colours of neighbor pixels *p* and *q*. The term $|X(p)-X(q)|$ makes

$E_2(X(p), X(q)) = 0$ whenever p and q belong to the same set.

The energy terms are used for defining costs for each edge in the graph. E_1 defines costs for a pixel belonging to a foreground (4) or background (5) region and E_2 set penalties when p and q nodes are assigned with different labels (6). The $\lambda = 1$ term is an empirical value.

Note that in this form there are no pixels marked as object or background in the image to be segmented. The energy terms were conceived to gather skin and non-skin information only from database images. Thus, the user presents an image that is segmented automatically based on the database colours.

The presented work in [1] are very similar to our, principally because it handles the segmentation with graph-cuts. However, the automatic selection of seeds from [1] is done by a K-means which limits the final result, different from our where choosen seeds are based on all images from database and limits our solution. Our method can handle different situations as shown in experimental results, but it can give very reliable results when pictures have the same illumination and similar poses.

V. EXPERIMENTAL RESULTS

The results presented in Figures 1, 3, and 4 were obtained with k_o = 128 and k_b = 128 mean colours. On the database, 24 images are used with skin and non-skin marked colours by a user. The images of the database are depicted in Figure 2. The images segmented in this section do not belong to the database only during the experiment, giving 23 images database on tests. When more images and consequently skin and non-skin colours are stored, more samples of different tones are considered, improving the average energy and obtaining better results. Experiments were evaluated using people with different skin tones. The algorithm used for computing the max-flow/min-cut in this paper is the Boykov-Kolmogorov [8].

Figure 1 shows an example of two images and their automatic human skin segmentation. Our method was applied to images of people of different ethnic groups. The results are consistent because the possible skin colours forms a small subset of the RGB space. Consequently, a skin pixel tends to have small distances for all images in the database, minimizing the cost in Equation (4).

The segmentation performs better when the database has a larger number of images. When more background pixels are marked, fewer potentially errors can occur in the segmentation of an image. This is shown in Figures 3(a), 3(b) and 3(c).

Figure 1. Segmentation of images that do not belong to the database.

Figure 2. Figures of the image database.

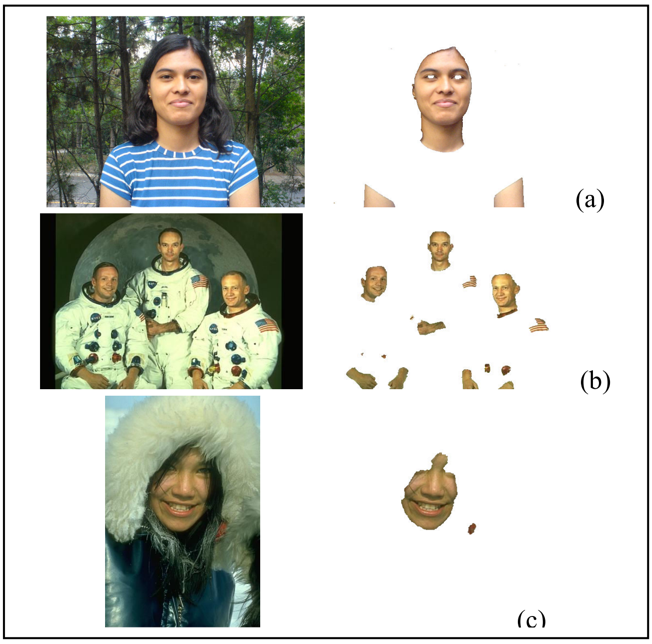

Figure 3. Example of segmentation with complex backgrounds.

However, segmentation may fail when an image has skin tones with colour coordinates very far from those of the corresponding colour cluster obtained from the database. This is possible because the colour clusters computed from the database may not cover the entire subset of RGB colour that corresponds to skin tones. Another problem occurs when the background has colours very similar to those present in skin regions (Figure 4). Even including the images with similar background is not enough to solve the problem. On Figure 4(a) the segmentation is performed without the image in database. Figure 4(b) shows the result when an image with adequate skin and non-skin seeds has been added. In these images it is possible to see the lines that the user draws, in order to indicate skin and non-skin regions. Note that this is not enough to produce good results for hair regions. In this case, colour based segmentation might not be enough.

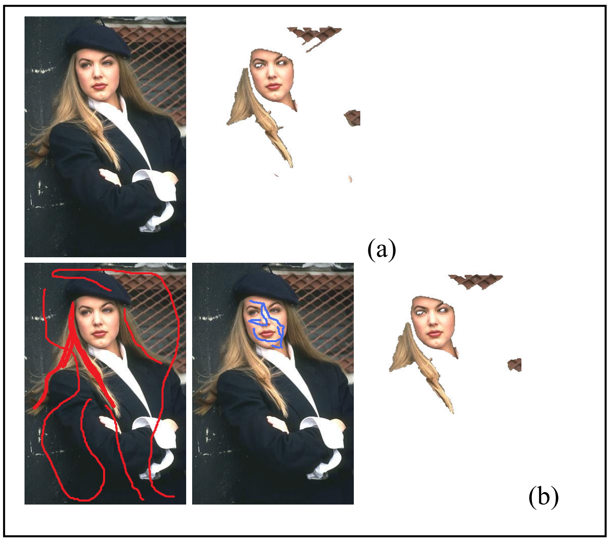

Figure 4. Adding an image with hair regions having colours close to skin tones.

In order to improve the results for generic images, we have built a database with heterogeneous images. However, the possibility of having background pixels with colours similar to skin pixels might lead to improper classification. A possible solution is to combine this method with other types of features as, for instance, textures.

VI. CONCLUSION

We present a new approach for automatic human skin segmentation. On traditional implementations of graph-cuts, identification of pixels is necessary for each image to be segmented, what in this work it is not needed. The method uses a database of marked images which gives clues for the algorithm on what regions are skin or non-skin for images to be segmented.

On future works we intend to use new colour spaces with graph-cuts approach as the HSL. The RGB color space is simple to use because of distance metric nature, but cannot treat variation principally because of illumination. Another possibility is to use a GPU parallel version of the min-cut/max-flow algorithm for video human skin segmentation.

VII. ACKNOWLEDGMENTS

Anselmo Montenegro is grateful for the fund from FAPERJ under process number E-26/171.208/2006. Marcelo Bernardes Vieira is thankful for the fund from FAPEMIG. Esteban Clua is grateful for the fund from FAPERJ. The others authors are grateful for CAPES.

REFERENCES

[1] Zhilan Hu, Guijin Wang, Xinggang Lin, Hong Yan, “Skin Segmentation Based on Graph Cuts,” Tsinghua Science & Technology, Volume 14, Issue 4, August 2009, Pages 478-486.

[2] L. Sigal, S. Sclaroff, and V. Athistos, “Estimation and prediction of evolving color distributions for skin segmentation under varying illumination,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 152-159, 2000.

[3] B. Jedynak, H. C. Zheng, and M. Daoudi. “Skin detection using pairwise models,” IVC(23), No. 13, 29 November 2005, pp. 1122-1130.

[4] M. Jones and J. Rehg, “Statistical color models with application to skin detection,” International Journal of Computer Vision, vol. 46, pp. 81-96, 2002.

[5] S. L. Phung, A. Bourzerdoum, and D. Chai, “Skin Segmentation using color and edge information,” Proc. Int. Symposium on Signal Processing and its Applications, 1-4, 2003.

[6] Y. Boykov and M. Jolly, “Interactive graph cuts for optimal boundary and region segmentation of n-d images,” Proc. IEEE Int. Conf. on Computer Vision, 105-112, 2001.

[7] Y. Li, J. Sun, C-K. Tang, and H.-Y. Shum, “Lazy Snapping,” ACM. Trans. Graph., vol. 23, no. 3, pp. 303-308, 2004.

[8] Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1124-1137, 2004.

[9] T. H. Cormen, C. E. Leiserson, and L. R. Rivest, Introduction to Algorithms. MIT Press and McGraw-Hill, 1990.

[10] L. R. Ford and A. D. R. Fulkerson, Flows in Networks. Princeton Univ. Pr. 1962.

[11] S. P. Lloyd. “Least square quantization in PCM,” IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129-137, March 1982.