Adaptive Detection of Human Skin in Color Images

Fernando Roberti de Siqueira
Institute of Computing
University of Campinas
Campinas, SP, Brazil, 13083-852

William Robson Schwartz
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte-MG, Brazil, 31270-010

Helio Pedrini
Institute of Computing
University of Campinas
Campinas, SP, Brazil, 13083-852

***Abstract*—Detection of human skin has several practical applications in image and video processing fields, such as face detection, image indexing, gesture recognition, and nudity detection. The purpose of a human skin detection method is to distinguish image portions between skin and non-skin regions. Many existing approaches in the literature are based on features that explore information of color, shape and texture in the images. Furthermore, several skin detection methods do not produce satisfactory results since they classify image regions based on fixed thresholds and are sensitive to illumination conditions. This paper describes a method for adaptive detection of human skin in images based on a normalized lookup table. The resulting probability map is used to detect skin and non-skin regions, which are refined through texture descriptors. Experimental results show the effectiveness of the proposed method.**

I. Introduction

Detection of human skin has applications in several areas, such as face recognition, gesture analysis, nudity detection, person tracking, content-based image retrieval, among others. The presence of people in an image or a video scene can be evidenced by finding skin regions.

Automatic skin detection is a challenging task, especially under varying illumination and partial occlusions [1]. Another inherent difficulty is that skin tones can significantly vary across individuals. Several methods found in the literature are based on fixed thresholds [2], [3] and are sensitive to geometric variations of skin patterns and are not robust to image resolution changes.

This paper describes and evaluates an adaptive human skin detection method based on a normalized lookup table, whose resulting probability map is used to detect skin and non-skin regions in the images. Such regions are then refined through texture descriptors. Experiments are conducted to apply the methodology to several images and results show the effectiveness of the proposed method.

The text is organized as follows. Section II presents the main concepts and work related to skin detection. Section III describes the proposed method for adaptive detection of human skin. Experimental results are shown in Section IV. Finally, Section V concludes the paper with final remarks.

II. Background

Several methods for skin detection have been proposed in the literature [4]. They can be categorized as pixel-based and region-based approaches. Pixel-based methods classify each pixel as skin or non-skin without considering its neighborhood, whereas region-based methods explore the spatial organization of neighbor skin pixels to improve the skin detection process. Comprehensive surveys on skin color modeling and skin detection can be found in the literature [4]–[6].

A skin color model is commonly used to identify if a pixel or region is skin or non-skin. Various color spaces have been used in skin detection, such as RGB [7], [8], HSV [9], CIE Lab [10], CIE Luv [11], YES [12] and YCbCr [2], [13].

Zarit et al. [14] compared five color spaces and two non-parametric methods for skin modeling. Shin et al. [15] examined eight color spaces applied to skin detection. Albiol [16] presented a theoretical proof that there is an optimum skin detector for every color space such that all skin detectors have the same performance. Brand and Mason [17] evaluated three skin color modeling strategies. Lee and Yoo [18] compared two most popular parametric skin models in different chrominance spaces.

Brown et al. [19] proposed a statistical model for skin color distribution based on a Self-Organizing Map (SOM). Sigal et al. [20] used a Markov model with adaptive color histogram to calculate the skin color distribution. Yang and Ahuja [11] described the use of expectation-maximization (EM) technique applied to Gaussian mixture models (GMM) for skin detection.

In contrast with methods described previously, the proposed method detects skin regions in the images through three consecutive stages, where each step is improved with the previous one. Measures of color, homogeneity and texture are taken into the account along the skin detection process.

III. Methodology

In this section, we present the process proposed to detect skin in images. The proposed approach is based not only of color information as traditional skin detection methods, but also on homogeneity and texture information to perform a more accurate detection.

First, we create a normalized lookup table (LUT) [7]. This is done by collecting measures of skin and non-skin pixel color samples and arranging them in a normalized histogram. This histogram provides a probability indicating how likely each pixel is skin or non-skin, such that a probability map is created to the entire image. With the application of a proper threshold, this map can be used to detect whether each pixel is skin or not.

To make the detection process more adaptable and achieve better results, a measure of how homogeneous is the detected region is evaluated since human skin regions tend to be more homogeneous than other types of surfaces [21]. Thus, the achieved results are maintained in an assessment of how homogeneous every region is according to the probability map, that is, regions that are not considered homogeneous are discarded, whereas homogeneous regions are grown as long as they remain homogeneous and then used in the output.

Finally, after considering two skin properties, color and homogeneity, the detection process is refined by taking texture information into account. Figure 1 shows the main stages of the proposed method.

A. Construction of Human Skin Color Model

To formulate the skin color model, we used patches of several images collected from the Internet. These patches were manually selected in skin and non-skin groups. Several non-skin samples were extracted from the Caltech image dataset [22].

Color histograms were constructed through both the skin and non-skin groups of images. RGB images were used to construct these histograms, each pixel forms a vector [RGB] which is translated to the lookup table as

$$H([RGB]) = H(R + [G * 256] + [B * 256 * 256]) \quad (1)$$

To find a probability of a pixel being in each group

$$P([RGB]) = \frac{H([RGB])}{\sum_{n=1}^{|H|} H(n)} \quad (2)$$

For every test image, we use a threshold t and the LUT to classify the image as follows [23]

$$\text{if } \frac{P(\text{pixel} \mid \text{skin})}{P(\text{pixel} \mid \neg \text{ skin})} \geq t \text{ then } P \text{ is labeled as skin} \quad (3)$$

where P(pixel | skin) is the probability of a pixel containing skin (informed by the histogram of the skin group) and P(pixel |¬ skin) is the probability of a pixel being in the non-skin group.

B. Selection of Homogeneous Regions

The LUT produces as an output a color map with the likelihood of every pixel being skin or not skin [21]. At this point, we select an initial threshold such that, if the probability of a point in the map is greater than the threshold, then it is labeled as skin, otherwise it is labeled as non-skin. Then, we create a black and white representation of the image (black being non-skin and white being skin). This gives us several skin-labeled connected components which will form a region. For each region of the image, we apply the following steps:

- if the region is smaller than a given threshold, it is discarded (the threshold size varies with the size of the image).
- while the region is not homogeneous, the threshold is increased by 10%.
- when a homogeneous region is found, we then dilate this selected image portion using a 3×3 square structure element as long as it remains homogeneous and is still smaller than the bounding box of the region.
- this homogeneous region is added to the skin list.

A region of the image is assigned as homogeneous if it has the following properties [21]:

$$(\sigma < S_t) \text{ AND } ((\frac{N_e}{N_d} \leq N_{dT}) \text{ OR } (\frac{N_e}{N_s} \leq N_{sT})) \quad (4)$$

where σ is the standard deviation of the pixel colors in the region, N_e is the number of edge pixels (found by using Sobel detector [24] in the entire input image) inside the region (boundary pixels are excluded), N_s the is number of skin pixels in the region and N_d is the maximum dimension of the region bounding box (either width or height, whichever is the largest). S_t, N_{dT} and N_{sT} are thresholds empirically chosen in the experiments, as detailed in Section IV.

C. Refinement with Texture Descriptors

After the previously described process has been applied to the images, we use a classifier trained to detect skin textures. The training was conducted on a Quadratic Discriminant Analysis (QDA) classifier [25] with the same images used to create the LUT histograms.

To perform the tests, we used a small window (32×32) sliding through the image, assigning how likely each window is of containing skin, creating a skin likelihood map. Then, we used a varying threshold to check the accuracy of the classification results.

IV. Experimental Results

The experiments were performed by applying the proposed methodology described in Section III. For the training of both normalized lookup table (LUT) [7] and quadratic discriminant analysis (QDA) classifier [25], we used randomly selected samples from the Internet and images extracted from the Caltech image dataset [22].

To form the homogeneity measure, the following variations of the parameters were employed: 35, 40 and 45 for S_t , 1.5, 2.5 and 3.5 for N_{dt} and 0.02, 0.12 and 0.22 for N_{st}. For the Sobel filter, we used the following thresholds: 20, 90, 160 and 230. We discarded any region that is smaller than a threshold in a range of 250, 500, 750 and 1000, varying with the size of the image (larger images used higher values).

The tests were conducted on the MCG skin-dataset [26], which contains 1000 images containing humans with skin exposure. The authors provide a ground truth for evaluation purposes. Each image was applied to the proposed method, producing an output that was compared against the ground truth. True positive and false positive rates were computed, such that the best combinations for each of the parameters were selected in a small subgroup of the original dataset, then applied to the full one. The values are shown in Table I.

Figure 2 illustrates the results by applying the proposed method to three different image samples. In the first row, it is possible to see that the result after applying the texture descriptors was not improved in relation the result obtained with skin color model and selection of homogeneous regions. On the other hand, in the second and third rows, the texture descriptors improved the result obtained in the previous stage, which increases the accuracy of the classifier.

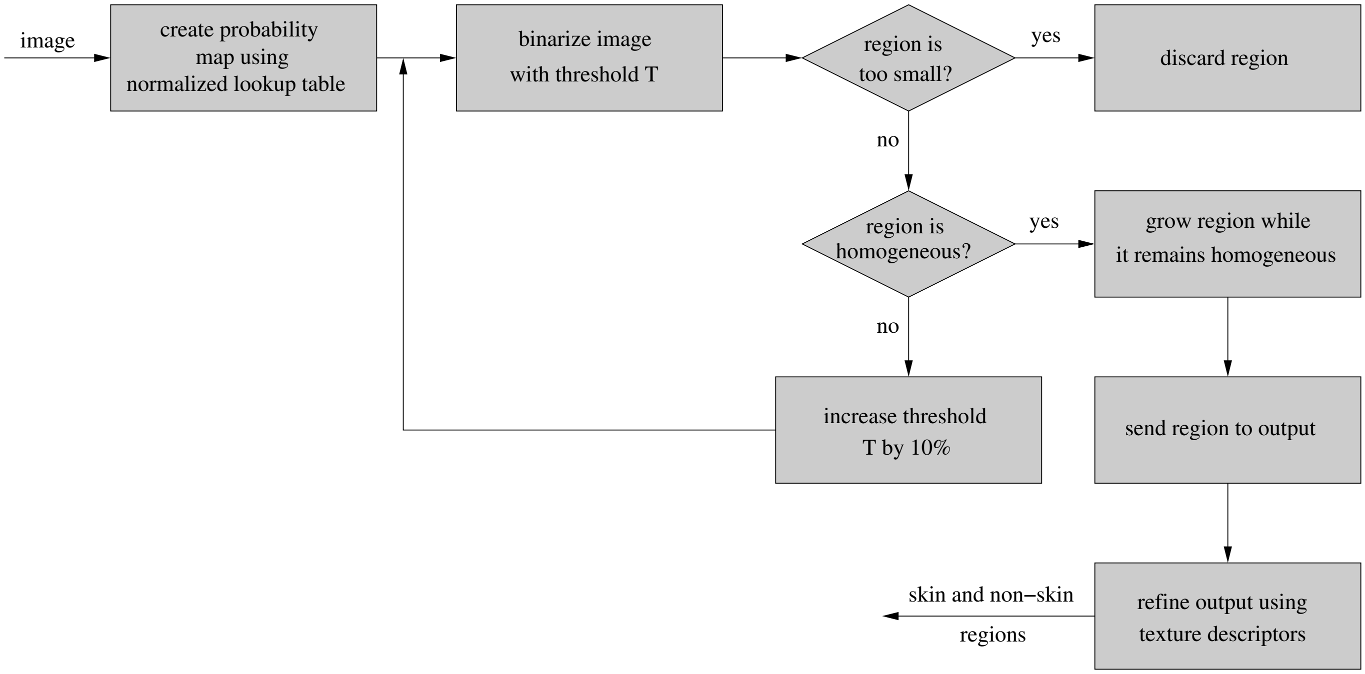


Figure 1. Diagram with the main stages of the skin detection process.

Textures	1	2	3	4	5	6
N_{dt}	1.5	2.5	3.5	1.5	2.5	3.5
N_{st}	0.22	0.22	0.22	0.02	0.02	0.02
Sobel threshold	20	20	20	90	90	90
S_t	40	40	45	45	45	40
True Positive Rate	67.14%	73.86%	77.81%	84.59%	85.50%	88.22%
False Positive Rate	13.08%	15.10%	16.78%	19.37%	20.62%	23.18%

Table I. VALUES OF THE MOST DISTINGUISHABLE PARAMETER COMBINATIONS.

Finally, to conduct the experiments with the texture classifier, the following parameters were used: $S_t = 40$, Sobel threshold = 90, $N_{st} = 0.02$ and $N_{dt} = 3.5$. This parameter combination produced the best true positive values, subjected to further refinement.

We assessed three different texture descriptors: Local Binary Patterns (LBP) [27], Grey Level Co-occurrence Matrices (GLCM) [28] and a Multiscale GLCM [29]. A threshold was used to build the likelihood map into skin and non-skin regions. This threshold was sampled between -30 and 7 (minimum and maximum likelihood, respectively), each sample providing a binary image compared to the ground truth to produce true and false positive measures, where several points were collected from each measure. Figure 3 shows a comparison among these textures detections along with the results obtained by not using any texture detection. According to the results, while the LBP feature descriptor achieved the lowest results, the remaining approaches showed similar performances.

V. CONCLUSIONS AND FUTURE WORK

This paper presented an adaptive human skin detection method based on a probability map used to detect skin and non-skin regions.

It was experimentally found that the texture stage has the potential to make some improvements on the results. However, in images that already presented high detection rate after the detection of homogeneous regions, the texture stage caused negative effect. The reason for that is probably due to the training process that was not sufficient to promote accurate distinction between region and non-regions. The training with a broader spectrum of skin images can improve the results.

As future directions, we intend to increase the number of training texture samples as an attempt to decrease the negative effect on some images and also improve the proposed method through a face detection process to adaptively enhance the skin color model to each image under different illumination conditions.

ACKNOWLEDGEMENTS

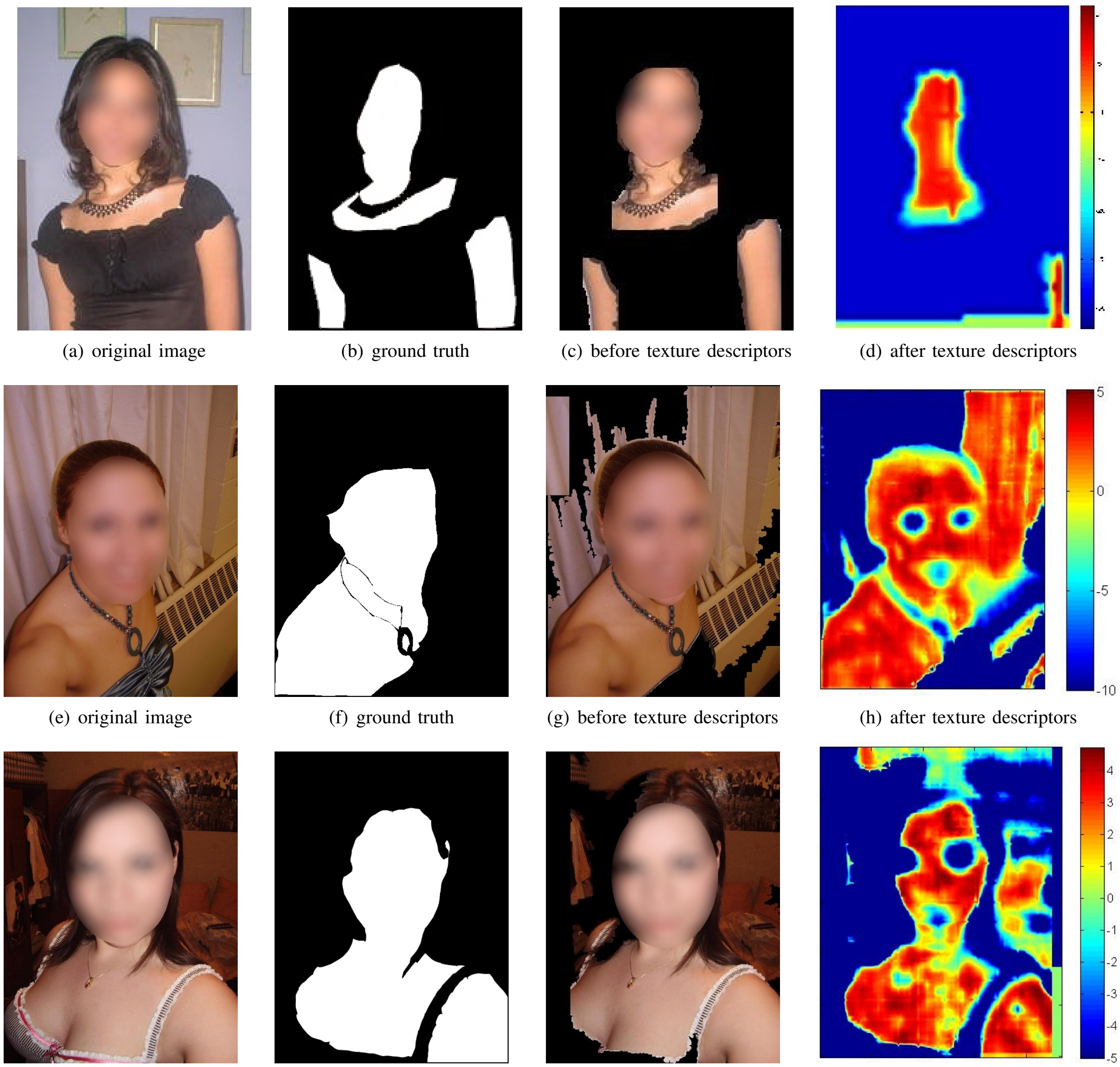
The authors are grateful to FAPESP, FAPEMIG, CNPq, and CAPES for the financial support.

REFERENCES

[1] S. Phung, A. Bouzerdoum, and D. Chai, “Skin segmentation using color pixel classification: analysis and comparison,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148–154, 2005.

[2] D. Chai and K. Ngan, “Face segmentation using skin-color map in videophone applications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551–564, Jun. 1999.

[3] C. Garcia and G. Tziritas, “Face detection using quantized skin color regions merging and wavelet packet analysis,” *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264–277, Sep. 1999.



(a) original image (b) ground truth (c) before texture descriptors (d) after texture descriptors

(e) original image (f) ground truth (g) before texture descriptors (h) after texture descriptors

(i) original image (j) ground truth (k) before texture descriptors (l) after texture descriptors

Figure 2. Results for two samples of images. (a), (e) and (i) original images; (b), (f) and (j) ground truth of the images; (c), (g) and (k) results after applying skin color model and selecting homogeneous regions; (d), (h) and (l) results after applying texture descriptors.

[4] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, “A survey of skin-color modeling and detection methods,” *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[5] V. Vezhnevets, V. Sazonov, and A. Andreeva, “A survey on pixel-based skin color detection techniques,” in *GraphiCon*, 2003, pp. 85–92.

[6] C. Prema and D. Manimegalai, “Survey on Skin Tone Detection using Color Spaces,” *International Journal of Applied Information Systems*, vol. 2, no. 2, pp. 18–26, May 2012.

[7] M. J. Jones and J. M. Rehg, “Statitical Color Models with Application to Skin Detection,” *Computer Vision*, vol. 46, pp. 81–86, 2002.

[8] G. Xu and T. Sugimoto, “Rits Eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera,” in *International Conference on Pattern Recognition*, Brisbane, Australia, Aug. 1998, pp. 1194–1197.

[9] N. Bojic and K. Pang, “Adaptive skin segmentation for head and shoulder video sequence,” in *SPIE Visual Communications and Image Processing*, Perth, Australia, Jun. 2000.

[10] J. Cai, A. Goshtasby, and C. Yu, “Detecting human faces in color images,” in *IEEE Internatinal Workshop on Multimedia Database Management Systems*, Austin, TX, USA, Aug. 1998, pp. 124–131.

[11] M.-H. Yang and N. Ahuja, “Gaussian mixture model for human skin color and its applications in image and video databases,” in *SPIE: Storage and Retrieval for Image and Video Databases*, San Jose, CA, USA, Jan. 1999.

[12] E. Saber and A. Tekalp, “Frontal-View Face Detection and Facial Feature Extraction using Color, Shape and Symmetry based Cost Functions,” *Pattern Recognition Letters*, vol. 19, no. 8, pp. 669–680, 1998.

[13] H. Wang and S. F. Chang, “A highly efficient system for automatic face detection in MPEG video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 4, pp. 615–628, Aug. 1997.

[14] B. Zarit, B. Super, and F. Quek, “Comparison of Five Color Models in Skin Pixel Classification,” in *International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, 1999, pp. 58–63.

[15] M. Shin, K. Chang, and L. V. Tsap, “Does colorspace transformation

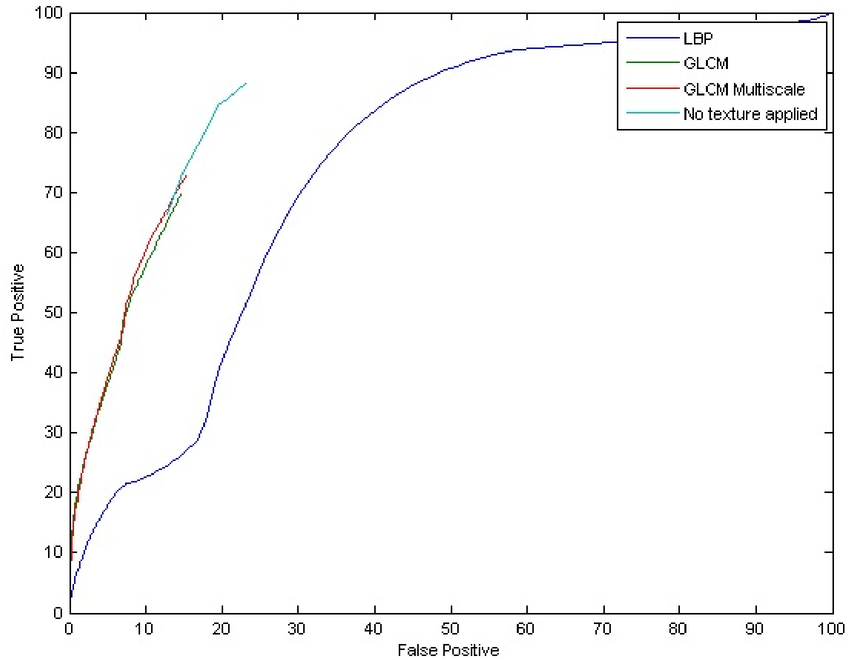


Figure 3. Results after applying different texture descriptors.

make any difference on skin detection?" in *Sixth IEEE Workshop on Applications of Computer Vision*, 2002, pp. 275–279.

[16] A. Albiol, L. Torres, and E. Delp, "Optimum color spaces for skin detection," in *International Conference on Image Processing*, vol. 1, Thessaloniki, Greece, Oct. 2001, pp. 122–124.

[17] J. Brand and J. Mason, "A Comparative Assessment of Three Approaches to Pixel-level Human Skin-Detection," in *International Conference on Pattern Recognition*, vol. 1, Barcelona, Spain, Sep. 2000, pp. 1056–1059.

[18] J. Lee and S. Yoo, "An elliptical boundary model for skin color detection," in *International Conference on Imaging Science, Systems and Technology*, Las Vegas, NV, USA, Jun. 2002.

[19] D. Brown, I. Craw, and J. Lewthwaite, "A SOM Based Approach to Skin Detection with Application in Real Time Systems," in *British Machine Vision Conference*, Manchester, United Kingdom, Sep. 2001.

[20] L. Sigal, S. Sclaroff, and V. Athitsos, "Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 582–587.

[21] S. L. Phung, D. Chai, and A. Bouzerdoum, "Adaptive skin segmentation in color images," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Hong Kong, China, Apr. 2003, pp. III–353–6.

[22] R. F. L. Fei-Fei and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories," *Workshop on Generative-Model Based Vision - IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2004.

[23] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, ser. Computer Science and Scientific Computing. Elsevier Science, 1990.

[24] R. Gonzalez, R. Woods, and S. L. Eddins, *Digital Image Processing using MATLAB*. Gatesmark Publishing, 2009.

[25] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed., ser. Springer Texts in Statistics. New York: Springer, 2005.

[26] L. Huang, T. Xia, Y. Zhang, and S. Lin, "Human skin detection in images by MSER analysis," in *18th IEEE International Conference on Image Processing*, Brussels, Belgium, 2011, pp. 1257–1260.

[27] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Featured Distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[28] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[29] F. Siqueira, W. Schwartz, and H. Pedrini, "Multi-Scale Gray Level Co-Occurrence Matrices for Texture Description," *Neurocomputing (to appear)*, 2013.